PURE AND APPLIED MATHEMATICS

A Series of Monographs and Textbooks

MATRICES OVER COMMUTATIVE RINGS

William C. Brown

PURE AND APPLIED MATHEMATICS

A Program of Monographs, Textbooks, and Lecture Notes

MONOGRAPHS AND TEXTBOOKS IN PURE AND APPLIED MATHEMATICS

1. *K. Yano,* Integral Formulas in Riemannian Geometry (1970)
2. *S. Kobayashi,* Hyperbolic Manifolds and Holomorphic Mappings (1970)
3. *V. S. Vladimirov,* Equations of Mathematical Physics (A. Jeffrey, ed.; A. Littlewood, trans.) (1970)
4. *B. N. Pshenichnyi,* Necessary Conditions for an Extremum (L. Neustadt, translation ed.; K. Makowski, trans.) (1971)
5. *L. Narici et al.,* Functional Analysis and Valuation Theory (1971)
6. *S. S. Passman,* Infinite Group Rings (1971)
7. *L. Dornhoff,* Group Representation Theory. Part A: Ordinary Representation Theory. Part B: Modular Representation Theory (1971, 1972)
8. *W. Boothby and G. L. Weiss, eds.,* Symmetric Spaces (1972)
9. *Y. Matsushima,* Differentiable Manifolds (E. T. Kobayashi, trans.) (1972)
10. *L. E. Ward, Jr.,* Topology (1972)
11. *A. Babakhanian,* Cohomological Methods in Group Theory (1972)
12. *R. Gilmer,* Multiplicative Ideal Theory (1972)
13. *J. Yeh,* Stochastic Processes and the Wiener Integral (1973)
14. *J. Barros-Neto,* Introduction to the Theory of Distributions (1973)
15. *R. Larsen,* Functional Analysis (1973)
16. *K. Yano and S. Ishihara,* Tangent and Cotangent Bundles (1973)
17. *C. Procesi,* Rings with Polynomial Identities (1973)
18. *R. Hermann,* Geometry, Physics, and Systems (1973)
19. *N. R. Wallach,* Harmonic Analysis on Homogeneous Spaces (1973)
20. *J. Dieudonné,* Introduction to the Theory of Formal Groups (1973)
21. *I. Vaisman,* Cohomology and Differential Forms (1973)
22. *B.-Y. Chen,* Geometry of Submanifolds (1973)
23. *M. Marcus,* Finite Dimensional Multilinear Algebra (in two parts) (1973, 1975)
24. *R. Larsen,* Banach Algebras (1973)
25. *R. O. Kujala and A. L. Vitter, eds.,* Value Distribution Theory: Part A; Part B: Deficit and Bezout Estimates by Wilhelm Stoll (1973)
26. *K. B. Stolarsky,* Algebraic Numbers and Diophantine Approximation (1974)
27. *A. R. Magid,* The Separable Galois Theory of Commutative Rings (1974)
28. *B. R. McDonald,* Finite Rings with Identity (1974)
29. *J. Satake,* Linear Algebra (S. Koh et al., trans.) (1975)
30. *J. S. Golan,* Localization of Noncommutative Rings (1975)
31. *G. Klambauer,* Mathematical Analysis (1975)
32. *M. K. Agoston,* Algebraic Topology (1976)
33. *K. R. Goodearl,* Ring Theory (1976)
34. *L. E. Mansfield,* Linear Algebra with Geometric Applications (1976)
35. *N. J. Pullman,* Matrix Theory and Its Applications (1976)
36. *B. R. McDonald,* Geometric Algebra Over Local Rings (1976)
37. *C. W. Groetsch,* Generalized Inverses of Linear Operators (1977)
38. *J. E. Kuczkowski and J. L. Gersting,* Abstract Algebra (1977)
39. *C. O. Christenson and W. L. Voxman,* Aspects of Topology (1977)
40. *M. Nagata,* Field Theory (1977)
41. *R. L. Long,* Algebraic Number Theory (1977)
42. *W. F. Pfeffer,* Integrals and Measures (1977)
43. *R. L. Wheeden and A. Zygmund,* Measure and Integral (1977)
44. *J. H. Curtiss,* Introduction to Functions of a Complex Variable (1978)
45. *K. Hrbacek and T. Jech,* Introduction to Set Theory (1978)
46. *W. S. Massey,* Homology and Cohomology Theory (1978)
47. *M. Marcus,* Introduction to Modern Algebra (1978)
48. *E. C. Young,* Vector and Tensor Analysis (1978)
49. *S. B. Nadler, Jr.,* Hyperspaces of Sets (1978)
50. *S. K. Segal,* Topics in Group Kings (1978)
51. *A. C. M. van Rooij,* Non-Archimedean Functional Analysis (1978)
52. *L. Corwin and R. Szczarba,* Calculus in Vector Spaces (1979)

53. *C. Sadosky,* Interpolation of Operators and Singular Integrals (1979)
54. *J. Cronin,* Differential Equations (1980)
55. *C. W. Groetsch,* Elements of Applicable Functional Analysis (1980)
56. *I. Vaisman,* Foundations of Three-Dimensional Euclidean Geometry (1980)
57. *H. I. Freedan,* Deterministic Mathematical Models in Population Ecology (1980)
58. *S. B. Chae,* Lebesgue Integration (1980)
59. *C. S. Rees et al.,* Theory and Applications of Fourier Analysis (1981)
60. *L. Nachbin,* Introduction to Functional Analysis (R. M. Aron, trans.) (1981)
61. *G. Orzech and M. Orzech,* Plane Algebraic Curves (1981)
62. *R. Johnsonbaugh and W. E. Pfaffenberger,* Foundations of Mathematical Analysis (1981)
63. *W. L. Voxman and R. H. Goetschal,* Advanced Calculus (1981)
64. *L. J. Corwin and R. H. Szczarba,* Multivariable Calculus (1982)
65. *V. I. Istrăţescu,* Introduction to Linear Operator Theory (1981)
66. *R. D. Järvinan,* Finite and Infinite Dimensional Linear Spaces (1981)
67. *J. K. Beem and P. E. Ehrlich,* Global Lorentzian Geometry (1981)
68. *D. L. Armacost,* The Structure of Locally Compact Abelian Groups (1981)
69. *J. W. Brewer and M. K. Smith, eds.,* Emily Noether: A Tribute (1981)
70. *K. H. Kim,* Boolean Matrix Theory and Applications (1982)
71. *T. W. Wieting,* The Mathematical Theory of Chromatic Plane Ornaments (1982)
72. *D. B. Gauld,* Differential Topology (1982)
73. *R. L. Faber,* Foundations of Euclidean and Non-Euclidean Geometry (1983)
74. *M. Carmeli,* Statistical Theory and Random Matrices (1983)
75. *J. H. Carruth et al.,* The Theory of Topological Semigroups (1983)
76. *R. L. Faber,* Differential Geometry and Relativity Theory (1983)
77. *S. Barnett,* Polynomials and Linear Control Systems (1983)
78. *G. Karpilovsky,* Commutative Group Algebras (1983)
79. *F. Van Oystaeyen and A. Verschoren,* Relative Invariants of Rings (1983)
80. *I. Vaisman,* A First Course in Differential Geometry (1984)
81. *G. W. Swan,* Applications of Optimal Control Theory in Biomedicine (1984)
82. *T. Petrie and J. D. Randall,* Transformation Groups on Manifolds (1984)
83. *K. Goebel and S. Reich,* Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings (1984)
84. *T. Albu and C. Năstăsescu,* Relative Finiteness in Module Theory (1984)
85. *K. Hrbacek and T. Jech,* Introduction to Set Theory: Second Edition (1984)
86. *F. Van Oystaeyen and A. Verschoren,* Relative Invariants of Rings (1984)
87. *B. R. McDonald,* Linear Algebra Over Commutative Rings (1984)
88. *M. Namba,* Geometry of Projective Algebraic Curves (1984)
89. *G. F. Webb,* Theory of Nonlinear Age-Dependent Population Dynamics (1985)
90. *M. R. Bremner et al.,* Tables of Dominant Weight Multiplicities for Representations of Simple Lie Algebras (1985)
91. *A. E. Fekete,* Real Linear Algebra (1985)
92. *S. B. Chae,* Holomorphy and Calculus in Normed Spaces (1985)
93. *A. J. Jerri,* Introduction to Integral Equations with Applications (1985)
94. *G. Karpilovsky,* Projective Representations of Finite Groups (1985)
95. *L. Narici and E. Beckenstein,* Topological Vector Spaces (1985)
96. *J. Weeks,* The Shape of Space (1985)
97. *P. R. Gribik and K. O. Kortanek,* Extremal Methods of Operations Research (1985)
98. *J.-A. Chao and W. A. Woyczynski, eds.,* Probability Theory and Harmonic Analysis (1986)
99. *G. D. Crown et al.,* Abstract Algebra (1986)
100. *J. H. Carruth et al.,* The Theory of Topological Semigroups, Volume 2 (1986)
101. *R. S. Doran and V. A. Belfi,* Characterizations of C*-Algebras (1986)
102. *M. W. Jeter,* Mathematical Programming (1986)
103. *M. Altman,* A Unified Theory of Nonlinear Operator and Evolution Equations with Applications (1986)
104. *A. Verschoren,* Relative Invariants of Sheaves (1987)
105. *R. A. Usmani,* Applied Linear Algebra (1987)
106. *P. Blass and J. Lang,* Zariski Surfaces and Differential Equations in Characteristic p > 0 (1987)
107. *J. A. Reneke et al.,* Structured Hereditary Systems (1987)

108. *H. Busemann and B. B. Phadke,* Spaces with Distinguished Geodesics (1987)
109. *R. Harte,* Invertibility and Singularity for Bounded Linear Operators (1988)
110. *G. S. Ladde et al.,* Oscillation Theory of Differential Equations with Deviating Arguments (1987)
111. *L. Dudkin et al.,* Iterative Aggregation Theory (1987)
112. T. Okubo, *Differential Geometry* (1987)
113. *D. L. Stancl and M. L. Stancl,* Real Analysis with Point-Set Topology (1987)
114. *T. C. Gard,* Introduction to Stochastic Differential Equations (1988)
115. *S. S. Abhyankar,* Enumerative Combinatorics of Young Tableaux (1988)
116. *H. Strade and R. Farnsteiner,* Modular Lie Algebras and Their Representations (1988)
117. *J. A. Huckaba,* Commutative Rings with Zero Divisors (1988)
118. *W. D. Wallis,* Combinatorial Designs (1988)
119. *W. Więsław,* Topological Fields (1988)
120. *G. Karpilovsky,* Field Theory (1988)
121. *S. Caenepeel and F. Van Oystaeyen,* Brauer Groups and the Cohomology of Graded Rings (1989)
122. *W. Kozlowski,* Modular Function Spaces (1988)
123. *E. Lowen-Colebunders,* Function Classes of Cauchy Continuous Maps (1989)
124. *M. Pavel,* Fundamentals of Pattern Recognition (1989)
125. *V. Lakshmikantham et al.,* Stability Analysis of Nonlinear Systems (1989)
126. *R. Sivaramakrishnan,* The Classical Theory of Arithmetic Functions (1989)
127. *N. A. Watson,* Parabolic Equations on an Infinite Strip (1989)
128. *K. J. Hastings,* Introduction to the Mathematics of Operations Research (1989)
129. *B. Fine,* Algebraic Theory of the Bianchi Groups (1989)
130. *D. N. Dikranjan et al.,* Topological Groups (1989)
131. *J. C. Morgan II,* Point Set Theory (1990)
132. *P. Biler and A. Witkowski,* Problems in Mathematical Analysis (1990)
133. *H. J. Sussmann,* Nonlinear Controllability and Optimal Control (1990)
134. *J.-P. Florens et al.,* Elements of Bayesian Statistics (1990)
135. *N. Shell,* Topological Fields and Near Valuations (1990)
136. *B. F. Doolin and C. F. Martin,* Introduction to Differential Geometry for Engineers (1990)
137. *S. S. Holland, Jr.,* Applied Analysis by the Hilbert Space Method (1990)
138. *J. Okniński,* Semigroup Algebras (1990)
139. *K. Zhu,* Operator Theory in Function Spaces (1990)
140. *G. B. Price,* An Introduction to Multicomplex Spaces and Functions (1991)
141. *R. B. Darst,* Introduction to Linear Programming (1991)
142. *P. L. Sachdev,* Nonlinear Ordinary Differential Equations and Their Applications (1991)
143. *T. Husain,* Orthogonal Schauder Bases (1991)
144. *J. Foran,* Fundamentals of Real Analysis (1991)
145. *W. C. Brown,* Matrices and Vector Spaces (1991)
146. *M. M. Rao and Z. D. Ren,* Theory of Orlicz Spaces (1991)
147. *J. S. Golan and T. Head,* Modules and the Structures of Rings (1991)
148. *C. Small,* Arithmetic of Finite Fields (1991)
149. *K. Yang,* Complex Algebraic Geometry (1991)
150. *D. G. Hoffman et al.,* Coding Theory (1991)
151. *M. O. González,* Classical Complex Analysis (1992)
152. *M. O. González,* Complex Analysis (1992)
153. *L. W. Baggett,* Functional Analysis (1992)
154. *M. Sniedovich,* Dynamic Programming (1992)
155. *R. P. Agarwal,* Difference Equations and Inequalities (1992)
156. *C. Brezinski,* Biorthogonality and Its Applications to Numerical Analysis (1992)
157. *C. Swartz,* An Introduction to Functional Analysis (1992)
158. *S. B. Nadler, Jr.,* Continuum Theory (1992)
159. *M. A. Al-Gwaiz,* Theory of Distributions (1992)
160. *E. Perry,* Geometry: Axiomatic Developments with Problem Solving (1992)
161. *E. Castillo and M. R. Ruiz-Cobo,* Functional Equations and Modelling in Science and Engineering (1992)
162. *A. J. Jerri,* Integral and Discrete Transforms with Applications and Error Analysis (1992)
163. *A. Charlier et al.,* Tensors and the Clifford Algebra (1992)

164. *P. Biler and T. Nadzieja*, Problems and Examples in Differential Equations (1992)
165. *E. Hansen*, Global Optimization Using Interval Analysis (1992)
166. *S. Guerre-Delabrière*, Classical Sequences in Banach Spaces (1992)
167. *Y. C. Wong*, Introductory Theory of Topological Vector Spaces (1992)
168. *S. H. Kulkarni and B. V. Limaye*, Real Function Algebras (1992)
169. *W. C. Brown*, Matrices Over Commutative Rings (1993)
170. *J. A. Loustau and M. I. Dillon*, Linear Geometry with Computer Graphics (1993)
171. *W. V. Petryshyn*, Approximation-Solvability of Nonlinear Functional and Differential Equations (1993)
172. *E. C. Young*, Vector and Tensor Analysis: Second Edition (1993)

Additional Volumes in Preparation

MATRICES OVER COMMUTATIVE RINGS

William C. Brown
Michigan State University
East Lansing, Michigan

Marcel Dekker, Inc. **New York • Basel • Hong Kong**

about the book . . .

This self-contained **reference/text** covers the **most important aspects** of the theory of matrices whose entries come from some given commutative ring—developing **all** the necessary facts about commutative rings throughout the body of the book.

Providing proofs that follow from concrete matrix calculations, ***Matrices Over Commutative Rings*** discusses the rank of a matrix, systems of linear equations, the Cayley-Hamilton theorem, resultants, Fitting ideals, and the Smith and Frobenius normal forms of matrices . . . studies linear algebra when the base ring is an arbitrary commutative ring, not necessarily a field . . . clarifies what the theory says when the base ring is a field, explaining the classical linear algebra results . . . and more.

Requiring only a general knowledge of abstract algebra, ***Matrices Over Commutative Rings*** serves as an excellent **reference** for algebraists; number theorists; and mathematicians, engineers, and scientists interested in matrix theory over commutative rings; and a long-needed **text** for graduate-level students in mathematics.

about the author . . .

WILLIAM C. BROWN is a Professor in the Department of Mathematics at Michigan State University, East Lansing. He is the author of over 30 research papers on noncommutative ring theory, commutative ring theory, nonassociative algebras, scheme theory, algebraic geometry, and linear algebra, plus three books on linear algebra, including *Matrices and Vector Spaces* (Marcel Dekker, Inc.). A member of the American Mathematical Society, Dr. Brown received the B.S. degree (1965) from the University of Miami, Coral Gables, Florida, and the Ph.D. degree (1969) from Northwestern University, Evanston, Illinois.

Printed in the United States of America ISBN: 0—8247—8755—2

marcel dekker, inc./new york · basel · hong kong

Library of Congress Cataloging-in-Publication Data

Brown, William C. (William Clough)
Matrices over commutative rings / William C. Brown
p. cm. — (Monographs and textbooks in pure and applied mathematics ; 169)
Includes bibliographical references and index.
ISBN 0-8247-8755-2 (acid-free paper)
1. Matrices. 2. Commutative rings. I. Title. II. Series.
QA188.B7598 1992
512.9'434—dc20 92-29098
CIP

This book is printed on acid-free paper.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

Preface

Matrices Over Commutative Rings is a textbook designed for a one-semester, graduate-level course in matrix theory. Any student who has studied linear algebra knows that there are many important situations in which the entries of a matrix are not field elements. The entries might be confined to the integers, or some principal ideal domain in general. The entries might be polynomials in finitely many variables or continuous functions on some finite interval. The general theory of matrices with entries from an arbitrary commutative ring is a vast subject and a very active research area in mathematics.

This text is not meant to be an encyclopedic description of the theory. In the 17 chapters of this book, I have discussed the topics that I feel are most important to every user of matrices. In particular, this text is not just for mathematicians. There are many topics here that engineers, business majors, scientists, and others will find interesting.

I have designed the book for a first-year, graduate level course in the subject. A general knowledge of abstract algebra is required to read this text. If the reader has had a good undergraduate course in abstract algebra, then he or she should be able to read this book with only an occasional detour to the appendices or references. I do not assume the reader has had a formal course in commutative ring theory. Most of the proofs are done from a matrix point of view; that is, certain matrix computations are performed and a result is obtained. In this way, the text can be read by a large audience, having no special background in

commutative ring theory. The necessary theorems in commutative ring theory are developed in the main body of the text as needed.

There is a good selection of problems at the end of each chapter. These problems range in difficulty from simple computations to proofs of hard theorems. Some of the exercises ask the reader to fill in missing details in arguments from the text. Although these types of problems may be boring to some, they often help further the reader's understanding of the subject matter. Some exercises involve the construction of complicated examples. I urge the reader to consult the general references at the end of this book when hunting for such examples. A glossary of notation has been provided for the reader's convenience at the end of this text.

There are four appendices at the end of this book. The first two deal with foundational material that is used sporadically throughout the rest of the text. Every student should know something about partially ordered sets and Zorn's lemma. Consequently, Appendix A should be read after Chapter 1 by every student using this book. Appendix B describes the Jacobson radical of a ring. This appendix will be of particular interest to students who wish to do advanced level work in algebra. Appendix B can also be read after Chapter 1, but, in general, the material found in Appendix B is a bit more difficult than the general level of the main text. Appendix C deals with applications of the theory of resultants. This appendix is also slightly more difficult than the main text. The reader should read Appendix C after Chapter 8. Appendix D presents an application of the ideas in Chapters 5 and 13. The interested reader should study Appendix D after Chapter 13.

Finally, I would like to say a few words about one of the general references at the end of this text. Neal McCoy's research in linear algebra over commutative rings was both important and influential. Many years ago, I first learned at least two of the topics in this text by reading McCoy's classic book *Rings and Ideals*. In particular, my treatment of the material in Chapters 6 and 9 is in the same spirit as McCoy's original treatment of these ideas. I would like to take this opportunity to acknowledge my indebtedness to McCoy's fine explanations of these topics.

William C. Brown

Contents

1

Modules Over Commutative Rings

All rings in this book will be associative and contain an identity (for multiplication), which we will usually denote by 1. If T is a ring, we will always assume $1 \neq 0$. Thus, T contains at least two distinct elements. We will let T^* denote the nonzero elements of T. Thus, $T^* = T - \{0\}$. An element $x \in T$ is called a left zero divisor of T if $xy = 0$ for some $y \in T^*$. The element x is called a right zero divisor if $yx = 0$ for some $y \in T^*$. We will let $Z(T)$ denote the set of elements in T which are either left or right zero divisors. Thus, if $x \in Z(T)$, then x is a left zero divisor of T or x is a right zero divisor of T. An element x in T is called a regular element if x is neither a left nor right zero divisor of T. Thus, $T - Z(T)$ is the set of regular elements in T. Notice that $0 \in Z(T)$, and 1 is a regular element of T.

In this book, we will always let R denote a commutative ring. For a commutative ring, the left zero divisors and right zero divisors are the same thing. Hence, $Z(R)$ will be called the set of zero divisors of R. If $x \in R - Z(R)$, then x is a regular element of R.

An element x in any ring T is called a unit if $xy = yx = 1$ for some $y \in T$. We will let $U(T)$ denote the units in T. Clearly, $U(T)$ is a group (the group operation being ring multiplication) contained in T^*. Two elements x and y in a commutative ring R are said to be associates if $x = uy$ for some unit $u \in U(R)$. If x and y are associates, we will write $x \sim y$. Notice that $\sim$ is an equivalence

relation on R. Thus, $x \sim x$, $x \sim y$ if and only if $y \sim x$ and if $x \sim y$ and $y \sim z$, then $x \sim z$.

Suppose R is a commutative ring and $x, y \in R$. The element x divides y if $xz = y$ for some $z \in R$. If x divides y, then we will write $x \mid y$. Notice that $x \mid y$ if and only if $Ry \subseteq Rx$. In particular, any element in R divides 0, and if $0 \mid y$, then $y = 0$. Let us consider a familiar example of some of these ideas.

Example 1.1 Let $T = M_{n\times n}(F)$ denote the ring of all $n \times n$ matrices with entries from a field F. We assume $n \geq 2$. Then T is a noncommutative ring. The units in T are the nonsingular matrices. Thus, $U(T) = \{A \in T \mid \det(A) \neq 0\}$. Here $\det(A)$ is the determinant of A. Suppose E_{ij} denotes the $n \times n$ matrix in T having a 1 in its i,jth entry and zeros elsewhere. Then each E_{ij} is both a right and left zero divisor in T. In fact, it is easy to check that $Z(T) = \{A \in T \mid \det(A) = 0\}$.

Let R denote the set of diagonal matrices in T. Then R is a commutative subring of T. A matrix in R is a unit in R if and only if no diagonal entry of R is zero. $Z(R)$ is the set of diagonal matrices for which at least one diagonal entry is zero. ∎

A commutative ring R is called an integral domain if $Z(R) = (0)$. Let us take this opportunity to introduce the notation we will use throughout the rest of this book for various familiar integral domains.

1.2 The following rings are all integral domains:

(a) $\mathbb{Z}$ = the ring of integers.
(b) F = an arbitrary field.
(c) $\mathbb{Q}$ = the field of rational numbers.
(d) $F[X_1, \ldots, X_n]$ = the ring of polynomials in variables $X_1, \ldots, X_n$ with coefficients from the field F.

All of the rings listed in 1.2 are integral domains with the usual definitions of addition and multiplication.

Let T be an arbitrary ring. The reader will recall that a nonempty subset $\mathfrak{A} \subseteq T$ is called a left ideal of T if $\mathfrak{A}$ is closed under addition and left multiplication. Thus, $\mathfrak{A}$ is a left ideal of T if $x + y \in \mathfrak{A}$ whenever $x, y \in \mathfrak{A}$, and $tx \in \mathfrak{A}$ whenever $x \in \mathfrak{A}$, and $t \in T$. Similarly, a nonempty subset $\mathfrak{B} \subseteq T$ is a right ideal of T if $x + y \in \mathfrak{B}$ whenever $x, y \in \mathfrak{B}$, and $xt \in \mathfrak{B}$ whenever $x \in \mathfrak{B}$, and $t \in T$. Notice that a left (or right) ideal of T is automatically a subgroup of the additive group $(T, +)$ of T.

A nonempty subset $\mathfrak{A}$ of T is a two-sided ideal of T if $\mathfrak{A}$ is both a left and right ideal of T. In this book, a two-sided ideal of T will be called an ideal of T. Thus, $\mathfrak{A} \subseteq T$ is an ideal of T if $\mathfrak{A}$ satisfies the following properties:

1.3 (a) $\mathfrak{A} \neq \varnothing$
(b) $x + y \in \mathfrak{A}$ for all $x, y \in \mathfrak{A}$
(c) $tx, xt \in \mathfrak{A}$ for all $x \in \mathfrak{A}$ and all $t \in T$

In 1.3a, the symbol $\varnothing$ denotes the empty set. Since $0 \neq 1$, T always contains at least two ideals, namely (0) and T.

If X and Y are subsets of some larger set Γ, we will use the notation $X < Y$ to mean X is a subset of Y and X is not equal to Y. Thus, $X < Y$ if and only if $X \subseteq Y$ and $X \neq Y$. A right, left, or two-sided ideal $\mathfrak{A}$ in a ring T is said to be proper if $\mathfrak{A} < T$. Thus, (0) is a proper ideal of T, while T is not a proper ideal of T. Consider the following simple example of left and right ideals.

Example 1.4 Let $T = M_{2\times 2}(\mathbb{Z})$ denote the set of 2×2 matrices with entries from $\mathbb{Z}$. The usual matrix operations in $M_{2\times 2}(\mathbb{Q})$ when restricted to T endow T with the structure of a noncommutative ring. The reader can easily verify the following statements:

(a) $\mathfrak{A} = \left\{ \begin{bmatrix} 0 & 0 \\ x & y \end{bmatrix} \;\middle|\; x,y \in \mathbb{Z} \right\}$ is a proper right ideal in T.

(b) $\mathfrak{B} = \left\{ \begin{bmatrix} x & 0 \\ y & 0 \end{bmatrix} \;\middle|\; x,y \in \mathbb{Z} \right\}$ is a proper left ideal in T.

(c) $\mathfrak{C} = \left\{ \begin{bmatrix} x & y \\ w & z \end{bmatrix} \;\middle|\; x,y,z,w \text{ even integers} \right\}$ is a proper ideal in T. ∎

Certainly, the most important example of an ideal in an arbitrary ring T is the Jacobson radical $J(T)$ of T. Suppose $\mathfrak{L}(T)$ denotes the set of proper left ideals of T. Notice that $\mathfrak{L}(T) \neq \varnothing$ since $(0) \in \mathfrak{L}(T)$. The elements of $\mathfrak{L}(T)$ can be partially ordered by inclusion $\subseteq$. A left ideal $\mathfrak{A} \in \mathfrak{L}(T)$ is maximal if $\mathfrak{A}$ is a maximal element in $\mathfrak{L}(T)$ with respect to the partial order $\subseteq$. Thus, $\mathfrak{A} \in \mathfrak{L}(T)$ is a maximal left ideal if, whenever $\mathfrak{A} \subseteq \mathfrak{B}$ and $\mathfrak{B} \in \mathfrak{L}(T)$, then $\mathfrak{A} = \mathfrak{B}$. The fact that $\mathfrak{L}(T)$ contains a maximal left ideal is a simple consequence of Zorn's lemma (see Example A.5 in Appendix A at the end of this text). The maximal left ideals in $\mathfrak{L}(T)$ are called maximal left ideals of T. The Jacobson radical of T is the intersection of all maximal left ideals of T. Thus, we have

1.5 $J(T) = \cap \{\mathfrak{A} \in \mathfrak{L}(T) \mid \mathfrak{A} \text{ is maximal}\}$.

There are a couple of places in the text where we will need various facts about $J(T)$. These facts are all summarized in Theorem 1.6. A proof of this theorem

and a complete discussion of the terminology appearing in Theorem 1.6 can be found at the end of this book in Appendix B.

Theorem 1.6 Let $J(T)$ denote the Jacobson radical of T.

(a) $J(T)$ is the intersection of all maximal left ideals of T.
(b) $J(T)$ is the intersection of all maximal right ideals of T.
(c) $J(T)$ is the intersection of all primitive ideals of T.
(d) $J(T)$ is a quasi-regular left ideal of T and contains every quasi-regular left ideal of T.
(e) $J(T)$ is a quasi-regular right ideal of T and contains every quasi-regular right ideal of T.
(f) $J(T) = \{z \in T \mid tz \text{ is left quasi-regular for all } t \in T\}$.
(g) $J(T) = \{z \in T \mid zt \text{ is right quasi-regular for all } t \in T\}$. ∎

Notice that Theorem 1.6b implies $J(T)$ is a proper ideal in T. If R is a commutative ring, then $\mathfrak{A}$ is a left ideal of R if and only if $\mathfrak{A}$ is a right ideal of R. Thus, the set of all left ideals of R, the set of all right ideals of R, and the set of all ideals of R are the same set. In this case, the Jacobson radical of R is just the intersection of all maximal ideals of R.

An ideal $\mathfrak{A}$ in a commutative ring R is said to be finitely generated if $\mathfrak{A} = Rx_1 + \cdots + Rx_n$ for some $x_1, \ldots, x_n \in \mathfrak{A}$. If every ideal in R is finitely generated, then R is called a Noetherian ring. We will have much more to say about Noetherian rings in later sections of this book. For now, we merely observe that the rings listed in 1.2 are all Noetherian rings. This is easy to see for $\mathbb{Z}$ or F. The fact that $F[X_1, \ldots, X_n]$ is Noetherian is a famous theorem in commutative ring theory called the Hilbert Basis Theorem.

A commutative ring R is called a principal ideal ring if every ideal $\mathfrak{A}$ of R is principal, that is, $\mathfrak{A} = Rx$ for some $x \in \mathfrak{A}$. In this book we will abbreviate the words principal ideal ring by writing PIR. Thus, R is a PIR if every ideal in R is generated by a single element. A PIR which is also an integral domain is called a principal ideal domain. We will abbreviate principal ideal domain by writting PID. Here is a list of some of the more well-known PIDs that students study in a first course in algebra.

1.7 The following rings are all PIDs:

(a) $\mathbb{Z}$
(b) $\mathbb{Z}[i]$ with $i^2 = -1$ (the Gaussian integers)
(c) $\mathbb{Z}[\sqrt{2}]$
(d) $\hat{\mathbb{Z}}_p$ (the p-adic integers)
(e) $F[X]$ (the polynomial ring in one variable over F)
(f) $F[[X]]$ (the power series ring in one variable over F)

Examples 1.7d and 1.7f may not be very familiar to the reader. See Exercises 10 and 11 at the end of this section for more details on these two examples.

Since the rings listed in Example 1.7 are all PIDs, any homomorphic image of any one of these rings is a PIR. In particular, $\mathbb{Z}/n\mathbb{Z}$ is a PIR for any integer $n \geq 2$. We will have more to say concerning the structure of PIRs in Chapter 14 of this book.

Now suppose R is a commutative ring. The reader will recall that an R-module M is an abelian group $(M, +)$ together with a function $f: R \times M \mapsto M$ (whose images we denote by $f(r, m) = rm$) which satisfies the following conditions:

1.8 $(x + y)m = xm + ym$ and $1m = m$
$x(m + n) = xm + xn$
$(xy)m = x(ym)$

These equations hold for all $x, y \in R$ and all $m, n \in M$. As we will see, R-modules are intimately related to the arithmetic of $M_{m \times n}(R)$, the set of $m \times n$ matrices with entries from R. We will need the usual definitions concerning R-module bases.

Definition 1.9 Let M be an R-module, and let $\Gamma = \{m_\alpha \mid \alpha \in \Delta\}$ be a subset of M.

(a) Γ is an R-module basis of M if every $m \in M$ can be written as a finite, linear combination of the elements of Γ.
(b) A finite subset $\{m_{\alpha(1)}, \ldots, m_{\alpha(n)}\}$ of distinct elements of Γ is said to be linearly independent over R if whenever $x_1 m_{\alpha(1)} + \cdots + x_n m_{\alpha(n)} = 0$ for some $x_1, \ldots, x_n \in R$, then $x_1 = \cdots = x_n = 0$.
(c) Γ is linearly independent over R if every finite subset of distinct elements from Γ is linearly independent over R.
(d) Γ is a free R-module basis of M if Γ is an R-module basis of M and Γ is linearly independent over R.
(e) M is a free R-module if M has a free R-module basis.

It might be wise to explain carefully some of the statements in Definition 1.9. The set Γ is an R-module basis of M if for every $m \in M$, there exist finitely many indices $\alpha(1), \ldots, \alpha(n) \in \Delta$ and ring elements $x_1, \ldots, x_n \in R$ such that $m = x_1 m_{\alpha(1)} + \cdots + x_n m_{\alpha(n)}$. The expression $x_1 m_{\alpha(1)} + \cdots + x_n m_{\alpha(n)}$ is called a linear combination of $m_{\alpha(1)}, \ldots, m_{\alpha(n)}$. We caution the reader that there is in general no uniqueness of representation here. It could be that m can be written in many different ways as linear combinations of elements from Γ. An R-module basis of M is often called a basis of M. In other words, the expression "R-module" is dropped. In particular, if Γ is a basis of M, this does not imply

every element of M can be written uniquely as a linear combination of elements from Γ. The use of the word basis is then different from what one encounters in vector space theory.

If $\Gamma = \{m_\alpha \mid \alpha \in \Delta\}$ is a free R-module basis of M, then every nonzero m in M can be written uniquely as a linear combination of elements from Γ. In other words, there exist unique indices $\alpha(1), \ldots, \alpha(n) \in \Delta$ and unique nonzero elements $x_1, \ldots, x_n \in R$ such that $m = x_1 m_{\alpha(1)} + \cdots + x_n m_{\alpha(n)}$. Of course, if $m = 0$, then $m = 0m_\alpha$ for any $\alpha \in \Delta$. Notice our definitions imply the zero module (0) is a free R-module with free R-module basis the empty set $\varnothing$.

Definition 1.10 An R-module M is said to be finitely generated if M has a finite basis $\Gamma = \{m_1, \ldots, m_n\}$.

The elements in a basis Γ are called generators of M. Thus, M is finitely generated if M has a finite set of generators. Consider the following examples.

Example 1.11

(a) Suppose V is a vector space over a field F. Any set of vectors which span V is a set of generators of the F-module V. Any vector space basis of V is a free F-module basis of V. Since every vector space has a vector space basis, every vector space is a free F-module. V is a finitely generated F-module precisely when $\dim_F(V) < \infty$.

(b) Let $M = R[X_1, \ldots, X_n]$, the polynomial ring over R in variables $X_1, \ldots, X_n$. Let Γ be the set of all monic monomials in $X_1, \ldots, X_n$. Thus

$$\Gamma = \{X_1^{\alpha(1)} X_2^{\alpha(2)} \cdots X_n^{\alpha(n)} \mid \alpha(1), \ldots, \alpha(n) \geq 0\}$$

Clearly, Γ is a free R-module basis of $R[X_1, \ldots, X_n]$. Notice that M is not a finitely generated R-module.

(c) Let

$$R^n = \{(x_1, \ldots, x_n)^t \mid x_1, \ldots, x_n \in R\}$$

Here $(x_1, \ldots, x_n)^t$ denotes the transpose of the row vector $(x_1, \ldots, x_n)$. Thus, R^n is the set of all column vectors of size n. With the usual definitions of addition and scalar multiplication, R^n is an R-module. Let $\varepsilon_1 = (1, 0, \ldots, 0)^t$, $\varepsilon_2 = (0, 1, 0, \ldots, 0)^t \ldots, \varepsilon_n = (0, \ldots, 0, 1)^t$. Then $\boldsymbol{\varepsilon} = \{\varepsilon_1, \ldots, \varepsilon_n\}$ is just the usual canonical basis of R^n. Clearly, $\boldsymbol{\varepsilon}$ is a free R-module basis of R^n. Therefore, R^n is a free, finitely generated R-module.

(d) The ring R itself is an R-module with $\{1\}$ being a free R-module basis of R.

(e) Let $R = \mathbb{Z}$, and suppose $n \geq 2$. Set $M = \mathbb{Z}/n\mathbb{Z}$. Then M is a $\mathbb{Z}$-module via the natural ring homomorphism from $\mathbb{Z}$ onto $\mathbb{Z}/n\mathbb{Z}$. If $\overline{1} = 1 + n\mathbb{Z}$ is the coset containing 1 in M, then $\{\overline{1}\}$ is a set of generators of M. Thus, $\mathbb{Z}/n\mathbb{Z}$ is a finitely generated $\mathbb{Z}$-module. Notice that $\mathbb{Z}/n\mathbb{Z}$ is not a free $\mathbb{Z}$-module since $n(\mathbb{Z}/n\mathbb{Z}) = (0)$. ∎

Definition 1.12 Let M be an R-module.

(a) For $m \in M$, $\text{Ann}_R(m) = \{x \in R \mid xm = 0\}$.

(b) $\text{Ann}_R(M) = \{x \in \text{R} \mid xm = 0 \text{ for all } m \in \text{M}\}$.

The set $\text{Ann}_R(m)$ is called the annihilator of m. Similarly, the set $\text{Ann}_R(M)$ is called the annihilator of M. Both $\text{Ann}_R(m)$ and $\text{Ann}_R(M)$ are ideals in R. Clearly, $\text{Ann}_R(M) = \cap \{\text{Ann}_R(m) \mid m \in M\}$. If M is a nonzero, free R-module, then $\text{Ann}_R(M) = (0)$. Thus, the first four examples in Example 1.11 (when $V \neq (0)$) have zero annihilators. On the other hand, $\text{Ann}_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z}) = n\mathbb{Z} \neq (0)$.

We can extend the notion of zero divisors to modules in an obvious way. An element $x \in R$ is a zero divisor of an R-module M if there exists a nonzero element $m \in M$ such that $xm = 0$. We will denote the set of zero divisors of M by $Z(M)$. In terms of annihilators, we have the following formula:

1.13 $Z(M) = \cup \{\text{Ann}_R(m) \mid m \in M - (0)\}$

We finish this section with some notation for R-module homomorphisms.

Definition 1.14 Suppose M and N are R-modules.

(a) A function $f : M \mapsto N$ is called an R-module homomorphism if $f(xm + ym') = xf(m) + yf(m')$ for all $m,m' \in M$ and $x,y \in R$.

(b) $\text{Hom}_R(M, N)$ will denote the set of all R-module homomorphisms from M to N.

(c) An R-module homomorphism f will be called an isomorphism if f is both one-to-one and onto.

(d) If $f: M \mapsto N$ is an isomorphism, then we will say M and N are isomorphic and write $M \cong N$.

(e) If $f \in \text{Hom}_R(M, N)$, then $\text{Ker}(f) = \{m \in M \mid f(m) = 0\}$ is called the kernel of f.

(f) If $f \in \text{Hom}_R(M, N)$, then $\text{Im}(f) = \{n \in N \mid n = f(m) \text{ for some } m \in M\}$ is called the image of f.

For any $f \in \text{Hom}_R(M, N)$, $\text{Ker}(f)$ is an R-submodule of M. Clearly, f is one-to-one if and only if $\text{Ker}(f) = (0)$. Similarly, $\text{Im}(f)$ is an R-submodule of N. The map f is onto if and only if $\text{Im}(f) = N$.

EXERCISES

1. Let $T = M_{n\times n}(F)$ as in Example 1.1. Show T has proper left and right nonzero ideals. Show T has no proper nonzero ideals.
2. In Exercise 1, show $Z(T) = \{A \in M_{n\times n}(F) \mid \det(A) = 0\}$.
3. Let $R = \mathbb{Z}/n\mathbb{Z}$. We assume $n \geq 2$. Compute the following sets:
   (a) $U(R)$
   (b) $Z(R)$
   (c) The ideals in R
   (d) $J(R)$
4. Let F be a field and set $R = F \times F \times \cdots \times F$. (There are n factors here with $n \geq 2$.) We define addition and multiplication componentwise. Show this is the commutative ring in Example 1.1.
5. For the ring R in Exercise 4, compute the following sets:
   (a) $U(R)$
   (b) $Z(R)$
   (c) The ideals in R
   (d) $J(R)$
6. Compute the units in the rings $\mathbb{Z}[\sqrt{10}]$ and $\mathbb{Z}[\sqrt{-5}]$ by using suitable multiplicative norms.
7. Let F be a field. Compute all ideals in F, and then show F is Noetherian.
8. Show that $\mathbb{Z}$ is a Noetherian ring.
9. For a commutative ring R, show $J(R) = \{x \in R \mid 1 + rx \in U(R) \text{ for all } r \in R\}$.
10. Let F be a field, and set $R = F[[X]] = \{\sum_{i=0}^{\infty} a_i X^i \mid a_i \in F\}$. R is the ring of formal power series over the field F. Addition and multiplication in R are defined as follows:

$$\left(\sum_{i=0}^{\infty} a_i X^i\right) + \left(\sum_{i=0}^{\infty} b_i X^i\right) = \sum_{i=0}^{\infty} (a_i + b_i) X^i$$

$$\left(\sum_{i=0}^{\infty} a_i X^i\right)\left(\sum_{i=0}^{\infty} b_i X^i\right) = \sum_{i=0}^{\infty} c_i X^i$$

Here $c_i = \sum_{p+q=i} a_p b_q$ for all $i \geq 0$.

   (a) Show that R is a commutative ring.
   (b) Determine $U(R)$.
   (c) Show every element in R can be written uniquely in the form δX^i for some $\delta \in U(R)$ and some $i \geq 0$.
   (d) Show $F[[X]]$ is a PID.

11. Let $R = \hat{\mathbb{Z}}_p = \{\sum_{i=0}^{\infty} a_i p^i \mid a_i \in \mathbb{Z}$, and $0 \leq a_i < p$ for all $i \geq 0\}$. Here p is a positive prime in $\mathbb{Z}$. Addition and multiplication of these series are computed as in the last exercise. When a coefficient is larger than p, the result is carried over to the next power of p. For example, when $p = 5$

$$(1 + 2.5^1 + 1.5^2 + \cdots) + (2 + 3.5^1 + 2.5^2 + \cdots)$$
$$= (3 + 0.5^1 + 4.5^2 + \cdots)$$

and

$$(1 + 2.5^1 + 1.5^2 + \cdots)(2 + 3.5^1 + 2.5^2 + \cdots)$$
$$= 2 + 2.5^1 + 1.5^2 + \cdots$$

(a) Show R is a commutative ring.
(b) Determine $U(R)$.
(c) Show every element in R can be written uniquely in the form δp^i for some $\delta \in U(R)$ and $i \geq 0$.
(d) Show $\hat{\mathbb{Z}}_p$ is a PID.

12. Show the integral domains in Exercise 6 are not PIDs.
13. Suppose R is an integral domain and $\mathfrak{A}$ is a proper ideal in R. Under what circumstances is $\mathfrak{A}$ a free R-module?
14. Let $\mathfrak{A}$ denote a proper nonzero ideal of R. Show the R-module $R/\mathfrak{A}$ is not a free R-module.
15. Is $\mathbb{Q}$ a free $\mathbb{Z}$-module?
16. Suppose M is a free R-module and N is a submodule of M. Then N need not be a free R-module. Construct a concrete example.
17. Suppose M is a direct sum of R-submodules $M_1, \ldots, M_n$. Thus, $M = M_1 \oplus \cdots \oplus M_n$. If each M_i is a free R-module, show M is also free.
18. The converse of Exercise 17 is not true. If M is a free R-module, then none of the M_i need be free. Use Exercise 4 to construct a concrete example.
19. Let M be an R-module. Set $\mathfrak{A} = \text{Ann}_R(M)$. Show M is an $R/\mathfrak{A}$-module in a natural way such that $\text{Ann}_{R/\mathfrak{A}}(M) = (0)$.
20. Let M and N be R-modules and suppose $M \cong N$. Thus, there exists an R-module homomorphism $f : M \mapsto N$ such that f is one-to-one and onto. Show there exists an R-module homomorphism $g : N \mapsto M$ which is one-to-one and onto. Thus, $M \cong N \Rightarrow N \cong M$.

2

Matrices with Entries from a Commutative Ring

In this chapter, we will review those aspects of matrix theory which are true over any commutative ring R.

Definition 2.1 The set of all $m \times n$ matrices with entries from R will be denoted by $M_{m\times n}(R)$.

If $A \in M_{m\times n}(R)$, we will let $[A]_{ij}$ denote the i,jth entry of A. The set $M_{m\times n}(R)$ is an R-module with addition and scalar multiplication defined in the usual ways. Thus, if $A,B \in M_{m\times n}(R)$, then $A + B$ is the $m \times n$ matrix whose i,jth entry is given by $[A + B]_{ij} = [A]_{ij} + [B]_{ij}$. If $r \in R$, then rA is the $m \times n$ matrix whose i,jth entry is given by $[rA]_{ij} = r[A]_{ij}$. It is a straightforward exercise to show $M_{m\times n}(R)$ is an R-module with these definitions of addition and scalar multiplication. The zero element in $M_{m\times n}(R)$ is the $m \times n$ matrix all of whose entries are zero. We will denote the zero matrix in $M_{m\times n}(R)$ by O. The size of the zero matrix will always be clear from the context in which it is being used.

For each $i = 1, \ldots, m$ and $j = 1, \ldots, n$, let E_{ij} denote the $m \times n$ matrix whose entries are defined as follows:

$$2.2 \quad [E_{ij}]_{pq} = \begin{cases} 1 & \text{if } (p,q) = (i,j) \\ 0 & \text{if } (p,q) \neq (i,j) \end{cases}$$

In equation 2.2, the index p ranges from 1 to m and q ranges from 1 to n. Thus, E_{ij} is the $m \times n$ matrix in $M_{m\times n}(R)$ which has a one in its i,jth entry and zeros elsewhere. The set of matrices $\Gamma = \{E_{ij} \mid 1 \leq i \leq m,\ 1 \leq j \leq n\}$ are called the matrix units of $M_{n\times n}(R)$ when $m = n$. At any rate, Γ is clearly a free R-module basis of $M_{m\times n}(R)$. Any matrix $A \in M_{m\times n}(R)$ can be written uniquely in terms of the E_{ij} as follows:

2.3 $A = \sum_{i=1}^{m} \sum_{j=1}^{n} [A]_{ij} E_{ij}$.

Thus, $M_{m\times n}(R)$ is a finitely generated, free R-module of rank mn. The rank of a free R-module is the number of generators in a free basis of the module. As we will see in Chapter 5, any two free bases of a finitely generated, free R-module M must have the same cardinality.

We will use the terms classical linear algebra and classical matrix theory to mean linear algebra and matrix theory over a field F. Any theorem in classical matrix theory whose proof requires only that F be a commutative ring is valid in $M_{m\times n}(R)$ for any commutative ring R. For example, the transpose A^t of a matrix A is defined by the usual formula: $[A^t]_{ij} = [A]_{ji}$. As in the classical case, the map $A \mapsto A^t$ is an R-module isomorphism from $M_{m\times n}(R)$ to $M_{n\times m}(R)$.

We do not intend to review all of classical matrix theory and make the corresponding translations to $M_{m\times n}(R)$. We will use any result in classical matrix theory which has an immediate translation to $M_{m\times n}(R)$ with little or no comment. However, there are a few results which we would like to emphasize in this chapter.

The usual theorems about block multiplication of partitioned matrices remain valid over any commutative ring. To understand these theorems, we need column and row vectors.

Definition 2.4 Let $A \in M_{m\times n}(R)$. The ith row of A will be denoted by $\text{Row}_i(A)$. The jth column of A will be denoted by $\text{Col}_j(A)$. Thus, if

$$A = \begin{bmatrix} a_{11}, \ldots, a_{1n} \\ \vdots \quad\quad \vdots \\ a_{m1}, \ldots, a_{mn} \end{bmatrix} \in M_{m\times n}(R)$$

we have the following formulas:

2.5 $\text{Row}_i(A) = (a_{i1}, a_{i2}, \ldots, a_{in})$ for $i = 1, \ldots, m$

$$\mathrm{Col}_j(A) = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \cdot \\ \cdot \\ \cdot \\ a_{mj} \end{bmatrix} = (a_{1j}, a_{2j}, \ldots, a_{mj})^t \quad \text{for } j = 1, \ldots, n$$

Notice that $\mathrm{Row}_i(A) \in M_{1\times n}(R)$ and $\mathrm{Col}_j(A) \in M_{m\times 1}(R)$. In keeping with the notation in most modern linear algebra texts, we set $R^m = M_{m\times 1}(R)$. Thus, R^m is the R-module consisting of all column vectors of size m. As we saw in Example 1.11c, R^m is a free R-module of rank m. We will adopt no special notation other than $M_{1\times n}(R)$ for row vectors of size n. If $A \in M_{m\times n}(R)$, then $\mathrm{Col}_j(A) \in R^m$ for $j = 1, \ldots, n$ and $\mathrm{Row}_i(A) \in M_{1\times n}(R)$ for $i = 1, \ldots, m$.

We can partition any $m \times n$ matrix A into m row vectors as follows:

$$\textbf{2.6} \quad A = \begin{bmatrix} \mathrm{Row}_1(A) \\ \hline \mathrm{Row}_2(A) \\ \hline \cdot \\ \cdot \\ \cdot \\ \hline \mathrm{Row}_m(A) \end{bmatrix}$$

The partition given in equation 2.6 is called the row partition of A. Henceforth, we will save space when writing this partition by using semicolons as follows:

$$\textbf{2.7} \quad A = (\mathrm{Row}_1(A); \mathrm{Row}_2(A); \ldots; \mathrm{Row}_m(A)).$$

We can further save space by designating each row of A by a Greek letter. If $\lambda_i = \mathrm{Row}_i(A)$ for $i = 1, \ldots, m$, then $A = (\lambda_1; \lambda_2; \ldots; \lambda_m)$.

We can also partition an $m \times n$ matrix A into n column vectors as follows:

$$\textbf{2.8} \quad A = (\mathrm{Col}_1(A) \mid \mathrm{Col}_2(A) \mid \ldots \mid \mathrm{Col}_n(A)).$$

The partition in equation 2.8 is called the column partition of A. If we set $\delta_j = \mathrm{Col}_j(A)$ for $j = 1, \ldots, n$, then $A = (\delta_1 \mid \delta_2 \mid \cdots \mid \delta_n)$.

If $A \in M_{m\times n}(R)$ and $B \in M_{n\times p}(R)$, then the product AB of A and B is defined in the usual way: $[AB]_{ij} = \sum_{k=1}^{n} [A]_{ik}[B]_{kj}$ for all $i = 1, \ldots m$ and $j = 1, \ldots, p$. The familiar theorem in classical matrix theory concerning the products of partitioned matrices is valid over any commutative ring.

Theorem 2.9 Let $A \in M_{m\times n}(R)$ and $B \in M_{n\times p}(R)$. Suppose

$$A = \left[\begin{array}{c|c|c|c} A_{11} & A_{12} & \cdots & A_{1k} \\ \hline \vdots & \vdots & & \vdots \\ \hline A_{r1} & A_{r2} & \cdots & A_{rk} \end{array}\right] \quad \text{and} \quad B = \left[\begin{array}{c|c|c|c} B_{11} & B_{12} & \cdots & B_{1t} \\ \hline \vdots & \vdots & & \vdots \\ \hline B_{k1} & B_{k2} & \cdots & B_{kt} \end{array}\right]$$

are partitions of A and B into submatrices such that each A_{ij} is an $m_i \times n_j$ matrix and each B_{jl} is an $n_j \times p_l$ matrix. Then $m_1 + m_2 + \cdots + m_r = m$, $n_1 + n_2 + \cdots + n_k = n$, and $p_1 + p_2 + \cdots p_t = p$. For each $i = 1, \ldots, r$ and $j = \ldots, t$, let $C_{ij} = \sum_{q=1}^{k} A_{iq} B_{qj}$. Then

$$AB = \left[\begin{array}{c|c|c|c} C_{11} & C_{12} & \cdots & C_{1t} \\ \hline \vdots & \vdots & & \vdots \\ \hline C_{r1} & C_{r2} & \cdots & C_{rt} \end{array}\right]$$

Proof. The proof of this result is the same as in the classical case. See [3, Chapter 1, Thm.3.10]. ∎

There are several special cases of Theorem 2.9 which are worth emphasizing here.

2.10 If $A \in M_{m\times n}(R)$, and $\xi = (x_1, \ldots, x_n)^t \in R^n$, then

$$A\xi = x_1 \operatorname{Col}_1(A) + x_2 \operatorname{Col}_2(A) + \cdots + x_n \operatorname{Col}_n(A)$$

Equation 2.10 is proved by partitioning A into columns and ξ into rows and applying Theorem 2.9. For any $A \in M_{m\times n}(R)$, we will let $\operatorname{CS}(A)$ denote the R-submodule of R^m generated by the columns of A. Equation 2.10 implies $\operatorname{CS}(A) = \{A\xi \mid \xi \in R^n\}$. The R-module $\operatorname{CS}(A)$ is called the column space of A.

Another application of Theorem 2.9 is the following result.

2.11 If $A \in M_{m\times n}(R)$ and $B \in M_{n\times p}(R)$, then

$$AB = (A \operatorname{Col}_1(B) \mid A \operatorname{Col}_2(B) \mid \cdots \mid A \operatorname{Col}_p(B))$$

Equation 2.11 is proved by leaving A alone, partitioning B into columns, and applying Theorem 2.9. Notice that equations 2.10 and 2.11 imply the jth column of AB is a linear combination of the columns of A with scalars from the jth column of B. In particular, we have the following familiar result:

2.12 If $A \in M_{m\times n}(R)$ and $B \in M_{n\times p}(R)$, then $CS(AB) \subseteq CS(A)$.

We also have the corresponding results for rows.

2.13 If $\delta = (x_1, \ldots, x_m) \in M_{1\times m}(R)$ and $A \in M_{m\times n}(R)$, then

$$\delta A = x_1 \operatorname{Row}_1(A) + x_2 \operatorname{Row}_2(A) + \cdots + x_m \operatorname{Row}_m(A)$$

For any $A \in M_{m\times n}(R)$, we will let $RS(A)$ denote the R-submodule of $M_{1\times n}(R)$ generated by the rows of A. The R-module $RS(A)$ is called the row space of A. Equation 2.13 implies $RS(A) = \{\delta A \mid \delta \in M_{1\times m}(R)\}$. We also have the analog of equation 2.11.

2.14 If $A \in M_{m\times n}(R)$ and $B \in M_{n\times p}(R)$, then

$$AB = (\operatorname{Row}_1(A)B; \operatorname{Row}_2(A)B; \ldots; \operatorname{Row}_m(A)B)$$

Equations 2.13 and 2.14 imply the ith row of AB is a linear combination of the rows of B with scalars from the ith row of A. Thus, we have the following analog of equation 2.12.

2.15 If $A \in M_{m\times n}(R)$ and $B \in M_{n\times p}(R)$, then $RS(AB) \subseteq RS(B)$.

The classical theory of determinants is another portion of linear algebra which is valid over any commutative ring R.

Definition 2.16 Let $A = (a_{ij}) \in M_{n\times n}(R)$. The determinant of A, written $\det(A)$, is the following element of R:

$$\det(A) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

In Definition 2.16, S_n denotes the set of all permutations on n letters. If $\sigma \in S_n$, then $\operatorname{sgn}(\sigma)$ denotes the sign of σ. Thus, $\operatorname{sgn}(\sigma) = 1$ if σ is even and $\operatorname{sgn}(\sigma) = -1$ if σ is odd. The sum in 2.16 is over all $n!$ permutations in S_n. This is just the classical definition of the determinant.

In the classical theory, the determinant is a multilinear function of its rows (or its columns). If $A,B \in M_{n\times n}(F)$, then $\det(AB) = \det(A)\det(B)$ and $\det(A^t) = \det(A)$. The proofs of these results use only the fact that F is a commutative ring. Hence, they are valid for any commutative ring R.

Let $A \in M_{m\times n}(R)$, and suppose $1 \le t \le \min\{m,n\}$. By a $t \times t$ minor of A, we mean the determinant of a $t \times t$ submatrix of A. Suppose Δ is a $t \times t$ minor of A. Then Δ is the determinant of a $t \times t$ submatrix of A. Now a $t \times t$ submatrix of A is chosen by selecting t rows and t columns of A. Suppose $i_1, \ldots, i_t$ are indices of the rows used to form Δ and $j_1, \ldots, j_t$ are indices of the columns used

to form Δ. We will usually assume $1 \leq i_1 < i_2 < \cdots < i_t \leq m$ and $1 \leq j_1 < j_2 < \cdots < j_t \leq n$. When we wish to specify which rows and columns are being used to compute Δ, we will write $\Delta(i_1, \ldots, i_t; j_1 \ldots j_t)$ in place of Δ. This notation is a bit cumbersome, but it has the redeeming feature of telling the reader exactly how Δ was formed. Consider the following example.

Example 2.17 Let

$$A = \begin{bmatrix} 1 & 2 & 3 & 3 \\ 1 & 0 & -1 & -1 \end{bmatrix} \in M_{2\times 4}(\mathbb{Z})$$

Then $\Delta(1;1) = 1$, $\Delta(2;3) = -1$

$$\Delta(1,2;1,2) = \det \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} = -2$$

and

$$\Delta(1,2;1,3) = \Delta(1,2;1,4) = \det \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix} = -4$$ ∎

Occasionally, we might want to abuse the definition and refer to $\Delta(\sigma(i_1), \ldots, \sigma(i_t); \tau(j_1), \ldots, \tau(j_t))$ where σ and τ are permutations of $i_1, \ldots, i_t$ and $j_1, \ldots, j_t$, respectively, as a minor of A. Since

$$\Delta(\sigma(i_1), \ldots, \sigma(i_t); \tau(j_1), \ldots, \tau(j_t)) = \pm \Delta(i_1 \ldots, i_t; j_1, \ldots, j_t)$$

this will cause no real confusion in the sequel. Notice that a given $m \times n$ matrix A has minors defined from submatrices of A of sizes $1 \times 1, 2 \times 2, \ldots, r \times r$ where $r = \min\{m,n\}$.

Definition 2.18 Let $A \in M_{n\times n}(R)$.

(a) For any $i,j = 1, \ldots, n$, $M_{ij}(A)$ will denote the $(n-1) \times (n-1)$ minor of A formed by deleting the ith row and jth column of A.

(b) The element $(-1)^{i+j}M_{ij}(A)$ will be called the i,jth cofactor of A. We will let $\text{cof}_{ij}(A)$ denote the i,jth cofactor of A.

In terms of our previous notation, we have the following expression for the i,jth cofactor of A:

$$\text{cof}_{ij}(A) = (-1)^{i+j}\Delta(1, \ldots, \hat{i}, \ldots, n; 1, \ldots, \hat{j}, \ldots, n)$$

Here the notation $\hat{i}$ and $\hat{j}$ means i and j are missing from the lists. Certainly the most important tool we will use from classical linear algebra is the Laplace expansion of the determinant. Laplace's theorem is valid over any commutative

ring R. Before stating this theorem, let us introduce one piece of useful notation. Throughout the rest of this book, we will let δ_{ij} denote the Kronecker delta function. Thus, $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. Laplace's theorem is as follows:

Theorem 2.19 (Laplace) Let $A = (a_{ij}) \in M_{n\times n}(R)$.
(a) $\sum_{j=1}^{n} a_{ij} \operatorname{cof}_{kj}(A) = \delta_{ik} \det(A)$ for all $i,k = 1, \ldots, n$.
(b) $\sum_{i=1}^{n} a_{ij} \operatorname{cof}_{ik}(A) = \delta_{jk} \det(A)$ for all $j,k = 1, \ldots, n$.

Proof. See [3, Chapter 3, Thm.3.13]. ∎

If we define the adjoint, $\operatorname{adj}(A)$, of an $n \times n$ matrix A by the usual formula:

$$[\operatorname{adj}(A)]_{ij} = \operatorname{cof}_{ji}(A) \quad \text{for all } i,j = 1, \ldots, n$$

then the formulas in Theorem 2.19 can be written succinctly as follows:

2.20 $A \operatorname{adj}(A) = \operatorname{adj}(A)\, A = \det(A)\, I_n$.

In equation 2.20, I_n denotes the $n \times n$ identity matrix. We can use equation 2.20 to determine the units in $M_{n\times n}(R)$, that is, the invertible matrices in $M_{n\times n}(R)$.

Corollary 2.21 Let $A \in M_{n\times n}(R)$. Then A is invertible if and only if $\det(A) \in U(R)$.

Proof. A is invertible if and only if $A \in U(M_{n\times n}(R))$. Suppose A is invertible. Then there exists a $B \in M_{n\times n}(R)$ such that $AB = BA = I_n$. Then

$$1 = \det(I_n) = \det(AB) = \det(A)\det(B)$$

In particular, $\det(A) \in U(R)$.

Conversely, if $\det(A) \in U(R)$, then equation 2.20 implies

$$A[(\det(A))^{-1} \operatorname{adj}(A)] = [(\det(A))^{-1} \operatorname{adj}(A)]A = I_n$$

Thus, A is invertible. ∎

The reader will notice that one classical result about determinants is no longer true for arbitrary commutative rings. If $\det(A) \neq 0$, we cannot conclude A is invertible.

Example 2.22 Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \in M_{2\times 2}(\mathbb{Z})$$

Then $\det(A) = 2$. Since 2 is not a unit in $\mathbb{Z}$, Corollary 2.21 implies A is not a unit in $M_{2\times 2}(\mathbb{Z})$. ∎

For any commutative ring R, we will let $\mathrm{Gl}(n,R)$ denote the set of invertible matrices in $M_{n\times n}(R)$. Since $\mathrm{Gl}(n,R) = U(M_{n\times n}(R))$, the set $\mathrm{Gl}(n,R)$ is a group with group operation matrix multiplication. The identity of this group is I_n. Corollary 2.21 implies

$$\mathrm{Gl}(n,R) = \{A \in M_{n\times n}(R) \mid \det(A) \in U(R)\}$$

$\mathrm{Gl}(n,R)$ is called the general linear group of $n \times n$ matrices over R.

We need one more definition before leaving this section. Diagonal matrices will play a central role in many of the theorems in this text.

Definition 2.23 Let $A \in M_{m\times n}(R)$. A is called a diagonal matrix if $[A]_{ij} = 0$ whenever $i \neq j$.

Notice that a diagonal matrix need not be square. Let $r = \min\{m,n\}$. We will use the notation $D = \mathrm{Diag}(d_1, \ldots, d_r)$ to denote an $m \times n$ diagonal matrix D whose $i,$ith entry is $[D]_{ii} = d_i$ for $i = 1, \ldots, r$. Thus, $\mathrm{Diag}(d_1, \ldots, d_r)$ has one of two forms depending on the order relationship between m and n.

2.24 If $m \leq n$, then $\mathrm{Diag}(d_1, \ldots, d_r)$ has the following form:

$$\mathrm{Diag}(d_1, \ldots, d_r) = \begin{bmatrix} d_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & & d_r & 0 & \cdots & 0 \end{bmatrix}.$$

If $m \geq n$, then $\mathrm{Diag}(d_1, \ldots, d_r)$ has the form

$$\mathrm{Diag}(d_1, \ldots, d_r) = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_r \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

The notation for a diagonal matrix does not include the size of the matrix. This will cause no confusion in what follows. The size of a diagonal matrix will always be clear from the context in which it appears.

EXERCISES

1. Show the ring $M_{n\times n}(\mathbb{Z})$ has infinitely many two-sided, proper ideals. Contrast this with the situation in Exercise 1 of Chapter 1.
2. Let $T = M_{n\times n}(R)$. Set

$$S = \{E_{ij} \mid 1 \leq i, j \leq n\} \cup \{0\}$$

Show S is closed under multiplication in T.

3. Suppose $R = R_1 \oplus \cdots \oplus R_s$ is a direct sum of (commutative) rings $R_1, \ldots, R_s$. Show

$$M_{n\times n}(R) \cong M_{n\times n}(R_1) \oplus \cdots \oplus M_{n\times n}(R_s)$$

4. Let T be a ring. The center, $C(T)$, of T is defined as follows:

$$C(T) = \{x \in T \mid xy = yx \text{ for all } y \in T\}$$

Compute $C(M_{n\times n}(R))$.

5. Let T be a ring. Let $M_{n\times n}(T)$ denote the set of all $n \times n$ matrices with entries from T. Define addition and multiplication in the usual ways: If $A,B \in M_{n\times n}(T)$, then $[A + B]_{ij} = [A]_{ij} + [B]_{ij}$ and $[AB]_{ij} = \sum_{k=1}^{n} [A]_{ik}[B]_{kj}$ for all $i,j = 1, \ldots n$. Show that $M_{n\times n}(T)$ is an associative ring with identity. Is $M_{n\times n}(T)$ a T-algebra ?
6. Show $M_{n\times n}(M_{m\times m}(R)) \cong M_{nm\times nm}(R)$ by using block addition and multiplication.
7. Let X be an indeterminate over the commutative ring R. Let $(M_{n\times n}(R))[X]$ denote the set of all polynomials in X with coefficients from $M_{n\times n}(R)$. Show that $(M_{n\times n}(R))[X]$ is a ring with the usual definitions of addition and multiplication of polynomials. Is $(M_{n\times n}(R))[X]$ commutative?
8. Using the notation in Exercise 7, show $(M_{n\times n}(R))[X] \cong M_{n\times n}(R[X])$.
9. Let $\mathrm{Tr} : M_{n\times n}(R) \mapsto R$ denote the trace mapping. Thus, if $A \in M_{n\times n}(R)$, then $\mathrm{Tr}(A) = \sum_{i=1}^{n} [A]_{ii}$. Verify the following statements:
   (a) $\mathrm{Tr} \in \mathrm{Hom}_R(M_{n\times n}(R),R)$.
   (b) $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ for all $A, B \in M_{n\times n}(R)$.
   (c) $\mathrm{Tr}(A^t) = \mathrm{Tr}(A)$ for all $A \in M_{n\times n}(R)$.
   (d) The R-submodule $\mathrm{Ker}(\mathrm{Tr})$ is generated as an R-module by the following set of matrices:

$$\{E_{ij} \mid 1 \leq i \neq j \leq n\} \cup \{E_{11} - E_{ii} \mid i \neq 1\}$$

10. Prove the assertions in 2.13 and 2.14.

11. Determine which of the following matrices are invertible in $M_{3\times 3}(\mathbb{Z})$ and compute their inverses.

(a) $\begin{bmatrix} -215 & -460 & -548 \\ 39 & 85 & 98 \\ 51 & 108 & 131 \end{bmatrix}$ (b) $\begin{bmatrix} 88 & 200 & 206 \\ -18 & -40 & -43 \\ -18 & -42 & -41 \end{bmatrix}$

(c) $\begin{bmatrix} 13 & -42 & 98 \\ -3 & 8 & -21 \\ -3 & 9 & -22 \end{bmatrix}$

12. If $AB = I_n$ in $M_{n\times n}(R)$, show $BA = I_n$.
13. Let

$$A = \left[\begin{array}{c|c|c|c} A_1 & O & \cdots & O \\ \hline O & A_2 & \cdots & O \\ \hline \vdots & \vdots & & \vdots \\ \hline O & O & \cdots & A_r \end{array}\right]$$

be a block diagonal matrix in $M_{n\times n}(R)$. We assume $A_i \in M_{n_i\times n_i}(R)$ for each $i = 1, \ldots, r$. Thus, $n_1 + \cdots + n_r = n$. Show $\det(A) = \prod_{i=1}^{r} \det(A_i)$. Therefore, A is invertible if and only if each A_i is invertible. Compute the inverse of A in terms of the inverses of the A_i.
14. Return to Exercise 3. Show $\mathrm{Gl}(n,R) \cong \mathrm{Gl}(n,R_1) \times \cdots \times \mathrm{Gl}(n,R_s)$.
15. Let $R = \mathbb{Z}/p\mathbb{Z}$ where p is a positive prime. Compute the order of the group $\mathrm{Gl}(n,\mathbb{Z}/p\mathbb{Z})$.
16. Let $\mathrm{Sl}(n,R) = \mathrm{Ker}(\det : \mathrm{Gl}(n,R) \mapsto U(R))$. Thus, $\mathrm{Sl}(n,R) = \{A \in M_{n\times n}(R) \mid \det(A) = 1\}$. Show $\mathrm{Sl}(n,R)$ is a normal subgroup of $\mathrm{Gl}(n,R)$. Compute the order of $\mathrm{Sl}(n,R)$ in Exercise 15. The group $\mathrm{Sl}(n,R)$ is called the special linear group over R.
17. Let $A \in M_{n\times n}(R)$, $\alpha \in R^n$, and $\beta \in M_{1\times n}(R)$. Show for any $z \in R$,

$$\det \left[\begin{array}{c|c} A & \alpha \\ \hline \beta & z \end{array}\right] = z \det(A) - \beta \operatorname{adj}(A)\, \alpha$$

3

The Ideals in $M_{n\times n}(R)$

If $\mathfrak{A}$ is an ideal of R, then clearly

$$M_{n\times n}(\mathfrak{A}) = \{A \in M_{n\times n}(R) \mid [A]_{ij} \in \mathfrak{A} \text{ for all } i,j = 1, \ldots, n\}$$

is an ideal (i.e., a two-sided ideal) of $M_{n\times n}(R)$. In fact, every ideal of $M_{n\times n}(R)$ is of this form.

Theorem 3.1 Let $\mathfrak{B}$ be an ideal of $M_{n\times n}(R)$. Then $\mathfrak{B} = M_{n\times n}(\mathfrak{A})$ for a unique ideal $\mathfrak{A}$ in R.

Before proving Theorem 3.1, we need to review some useful facts about matrix units. Let $\Gamma = \{E_{ij} \mid 1 \leq i,j \leq n\}$. be the matrix units introduced in equation 2.2. We have seen in Chapter 2 that Γ is a free R-module basis of $M_{n\times n}(R)$. We will need the following multiplication formulas:

3.2 (a) $E_{ij}E_{kl} = \begin{cases} E_{il} & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}$ for all $i,j,k,l = 1, \ldots, n$

(b) If $A = (a_{ij}) \in M_{n\times n}(R)$, then for all $p,q = 1, \ldots, n$,

$$AE_{pq} = \begin{bmatrix} 0, \ldots, 0, & a_{1p}, & 0, \ldots, 0 \\ 0, \ldots, 0, & a_{2p}, & 0, \ldots, 0 \\ \vdots & \vdots & \vdots \\ 0, \ldots, 0, & a_{np}, & 0, \ldots, 0 \end{bmatrix}$$
$$q$$

(c) If $A = (a_{ij}) \in M_{n\times n}(R)$, then for all $p,q = 1, \ldots n$,

$$E_{pq}A = \begin{bmatrix} 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \\ a_{q1}, & a_{q2}, \ldots, & a_{qn} \\ 0 & 0 \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 \cdots & 0 \end{bmatrix} \quad p$$

(d) If $A = (a_{ij}) \in M_{n\times n}(R)$, then for all $k,s,p,q = 1, \ldots, n$,

$E_{ks}AE_{pq} = a_{sp}E_{kq}$.

The formula in 3.2a follows directly from the definitions of the E_{ij} and matrix multiplication.

The q at the bottom of the matrix AE_{pq} in 3.2b means $(a_{1p}, \ldots, a_{np})^t$ is the qth column of AE_{pq}. Thus, a column partition of AE_{pq} is as follows:

$$AE_{pq} = (O \mid \cdots \mid O \mid \mathrm{Col}_p(A) \mid O \mid \cdots \mid O)$$

with $\mathrm{Col}_p(A)$ in the qth column of this matrix. To prove 3.2b, we have

$$\begin{aligned} AE_{pq} &= \left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}E_{ij}\right)E_{pq} \\ &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}E_{ij}E_{pq}\right) = \sum_{i=1}^{n} a_{ip}E_{iq} \\ &= (O \mid \cdots \mid O \mid \mathrm{Col}_p(A) \mid O \mid \cdots \mid O) \end{aligned}$$

with $\mathrm{Col}_p(A)$ in the qth column of this matrix.

The p at the right in 3.2c means the row $(a_{q1}, \ldots, a_{qn})$ is in row p of $E_{pq}A$. We have

$$E_{pq}A = E_{pq}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}E_{ij}\right)$$

$$= \sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}E_{pq}E_{ij}\right) = \sum_{j=1}^{n} a_{qj}E_{pj}$$

$$= (\mathrm{O}; \ldots; \mathrm{O}; \mathrm{Row}_q(A); \mathrm{O}; \ldots; \mathrm{O})$$

with $\mathrm{Row}_q(A)$ in row p of this matrix.

The proof of 3.2d follows from 3.2b. We have

$$E_{ks}AE_{pq} = E_{ks}\left(\sum_{i=1}^{n} a_{ip}E_{iq}\right) = \sum_{i=1}^{n} a_{ip}E_{ks}E_{iq} = a_{sp}E_{kq}$$

Thus, $E_{ks}AE_{pq}$ is the $n \times n$ matrix having a_{sp} in its k,qth entry and zeros elsewhere.

Proof of Theorem 3.1. Let $\mathfrak{B}$ be an ideal in $M_{n\times n}(R)$. Consider the set $\mathfrak{A}$ defined as follows:

$$\mathfrak{A} = \{r \in R \mid r \text{ is an entry in some matrix of } \mathfrak{B}\}$$

We will show the following statements are true:
(1) $\mathfrak{A}$ is an ideal of R.
(2) $\mathfrak{B} = M_{n\times n}(\mathfrak{A})$.
(3) If $\mathfrak{B} = M_{n\times n}(\mathfrak{C})$ for some ideal $\mathfrak{C}$ in R, then $\mathfrak{A} = \mathfrak{C}$.

We begin with (1). Clearly, $0 \in \mathfrak{A}$. Suppose $a,b \in \mathfrak{A}$. Then there exist two matrices $A, B \in \mathfrak{B}$ such that a is an entry in A and b is an entry in B, respectively. Suppose $a = [A]_{kp}$ and $b = [B]_{sq}$. Here k, p, s, and q are indices between 1 and n. From 3.2b, we have $[AE_{pq}]_{kq} = a$. From 3.2c, we have $[E_{ks}B]_{kq} = b$. Since $\mathfrak{B}$ is an ideal of $M_{n\times n}(R)$, $AE_{pq} \pm E_{ks}B \in \mathfrak{B}$. Thus, $a \pm b = [AE_{pq} \pm E_{ks}B]_{kq} \in \mathfrak{A}$ Since a, b are arbitrary elements of $\mathfrak{A}$, we conclude $\mathfrak{A}$ is an additive subgroup of R.

Now let $r \in R$ and $a \in \mathfrak{A}$. Then $a = [A]_{kp}$ for some $A \in \mathfrak{B}$ and some indices p, q between 1 and n. Since $\mathfrak{B}$ is an ideal in $M_{n\times n}(R)$, $\mathrm{Diag}(r, \ldots, r)A \in \mathfrak{B}$. In particular, $ra = [\mathrm{Diag}(r, \ldots, r)A]_{kp} \in \mathfrak{A}$. Thus, $\mathfrak{A}$ is an ideal in R and the proof of (1) is complete.

Clearly, $\mathfrak{B} \subseteq M_{n\times n}(\mathfrak{A})$. Suppose $A = (a_{ij}) \in M_{n\times n}(\mathfrak{A})$. Then $a_{ij} \in \mathfrak{A}$ for all $i,j = 1, \ldots, n$, and $A = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}E_{ij}$. We claim each $a_{ij}E_{ij}$ is a matrix in $\mathfrak{B}$. To see this, fix $i,j \in \{1, \ldots, n\}$. Since $a_{ij} \in \mathfrak{A}$, there exists a matrix B

$\in \mathfrak{B}$ such that $a_{ij} = [B]_{pq}$. Here p and q are fixed integers with $1 \leq p,q \leq n$. Using 3.2d, we have $a_{ij}E_{ij} = [B]_{pq}E_{ij} = E_{ip}BE_{qj} \in \mathfrak{B}$. Since A is the sum of the $a_{ij}E_{ij}$, we conclude $A \in \mathfrak{B}$. Thus, $M_{n\times n}(\mathfrak{A}) \subseteq \mathfrak{B}$, and the proof of (2) is complete.

As for (3), if $\mathfrak{B} = M_{n\times n}(\mathfrak{C})$, then clearly $\mathfrak{C} \subseteq \mathfrak{A}$. Let $a \in \mathfrak{A}$. Then $a = [B]_{pq}$ for some $B \in \mathfrak{B}$ and some $p,q \in \{1, \ldots, n\}$. Then $aE_{11} = E_{1p}BE_{q1} \in \mathfrak{B} = M_{n\times n}(\mathfrak{C})$. Therefore, $a \in \mathfrak{C}$. This completes the proof of (3).

The statements in (1), (2), and (3) obviously give a proof of Theorem 3.1. ∎

Theorem 3.1 implies the map $\mathfrak{A} \mapsto M_{n\times n}(\mathfrak{A})$ is a bijection from $\mathfrak{I}(R)$, the ideals of R, onto $\mathfrak{I}(M_{n\times n}(R)$, the ideals of $M_{n\times n}(R)$. The reader can easily check that this map satisfies the following properties.

3.3 Let $\mathfrak{A}_1, \mathfrak{A}_2 \in \mathfrak{I}(R)$.
(a) If $\mathfrak{A}_1 \subseteq \mathfrak{A}_2$ in $\mathfrak{I}(R)$, then $M_{n\times n}(\mathfrak{A}_1) \subseteq M_{n\times n}(\mathfrak{A}_2)$.
(b) $M_{n\times n}(\mathfrak{A}_1 \cap \mathfrak{A}_2) = M_{n\times n}(\mathfrak{A}_1) \cap M_{n\times n}(\mathfrak{A}_2)$.
(c) $M_{n\times n}(\mathfrak{A}_1 + \mathfrak{A}_2) = M_{n\times n}(\mathfrak{A}_1) + M_{n\times n}(\mathfrak{A}_2)$.
(d) $M_{n\times n}(\mathfrak{A}_1\mathfrak{A}_2) = M_{n\times n}(\mathfrak{A}_1)\, M_{n\times n}(\mathfrak{A}_2)$.

In other words, the map $\mathfrak{A} \mapsto M_{n\times n}(\mathfrak{A})$ preserves order, intersections, sums, and products as a function from the lattice of ideals of R to the lattice of ideals of $M_{n\times n}(R)$.

Suppose $R = F$, a field. Then $\mathfrak{I}(F) = \{(0),F\}$. Theorem 3.1 implies $\mathfrak{I}(M_{n\times n}(F)) = \{(0), M_{n\times n}(F)\}$. A ring T which has only two ideals, (0) and T, is called a simple ring. Hence, we have the following corollary to Theorem 3.1.

Corollary 3.4 If F is a field, then $M_{n\times n}(F)$ is a simple ring. ∎

Theorem 3.1 imples there is a one-to-one correspondence between the ideals of R and the ideals of $M_{n\times n}(R)$. The left and right ideals of $M_{n\times n}(R)$ do not come from R in any natural way. If R is a field, for example, then the only left (or right) ideals in R are (0) and R. However, $M_{n\times n}(R)$ has many proper left (or right) ideals. If $A \in M_{n\times n}(R)^*$ and $\det(A) = 0$, then $M_{n\times n}(R)A$ is a proper left ideal of $M_{n\times n}(R)$ and $AM_{n\times n}(R)$ is a proper right ideal of $M_{n\times n}(R)$.

We can use Theorem 3.1 to determine the Jacobson radical of $M_{n\times n}(R)$. Before stating the result, we need the two-sided Peirce decomposition of an element. Suppose T is a ring. Recall an element $e \in T$ is idempotent if $e^2 = e$. Suppose e is an idempotent element in T. Then for any $a \in T$, we have

3.5 $a = eae + ea(1 - e) + (1 - e)ae + (1 - e)a(1 - e)$.

The reader can easily check that the equality in 3.5 is true by multiplying out the right-hand side of the equation. The expression in equation 3.5 is called the two-sided Peirce decomposition of a (with respect to e). This expression will be used in the following lemma.

Lemma 3.6 Let T be a ring and e an idempotent element in T. Then $J(eTe) = eJ(T)e$

Proof. Before beginning the proof of this lemma, let us say a few words about what it means. If T is a ring, then eTe is also a ring with identity e. Hence, eTe has a Jacobson radical $J(eTe)$. On the other hand, $eJ(T)e$ is clearly an ideal in the ring eTe. The lemma asserts that these two ideals are the same in eTe.

The containment from right to left in $J(eTe) = eJ(T)e$ is easy. Suppose $x \in J(T)$. Then $ex \in J(T)$. Consequently, ex is left quasi-regular in T. Hence, $y(1 - ex) = 1$ for some $y \in T$. Then $ey(1 - ex)e = e$, or $(eye)(e - exe) = e$. Since e is the identity in eTe, the last equation implies exe is left quasi-regular in eTe. A similar proof shows exe is right quasi-regular in eTe. By Theorem 1.6, the Jacobson radical of eTe contains all quasi-regular left (or right) ideals of eTe. Therefore, $eJ(T)e \subseteq J(eTe)$.

For the other inclusion, let $z \in J(eTe)$. Let $a \in T$. From the Peirce decomposition of a, we have

$$\mathbf{3.7} \quad za = zeae + zea(1 - e) + z(1 - e)ae + z(1 - e)a(1 - e).$$

But $z \in eTe$. Therefore, $z(1 - e) = 0$. Hence, equation 3.7 becomes

$$\mathbf{3.8} \quad za = zeae + zea(1 - e).$$

Since $z \in J(eTe)$, $zeae \in J(eTe)$. Therefore, $zeae$ has a quasi-inverse $z' \in eTe$. Thus, $zeae \circ z' = z' \circ zeae = 0$ in eTe. [Here $x \circ y = x + y - xy$ is the circle composition (Appendix B).] Notice $zeae \circ z' = z' \circ zeae = 0$ in T since $eTe \subseteq T$. Thus, $zeae$ is a quasi-regular element of T as well as eTe. From equation 3.8, we have

$$\begin{aligned} za \circ z' &= [zeae + zea(1 - e)] \circ z' \\ &= zeae + zea(1 - e) + z' - (zeae + zea(1 - e))z' \\ &= zeae \circ z' + zea(1 - e) - zea(1 - e)z' = zea(1 - e) \end{aligned}$$

In this last equation, $zeae \circ z' = 0$ since z' is the quasi-inverse of $zeae$, and $zea(1 - e)z' = 0$ since $z' \in eTe$. Therefore, we have

$$\mathbf{3.9} \quad za \circ z' = zea(1 - e).$$

Since $z \in eTe$, the element $zea(1 - e)$ is nilpotent of index ≤ 2. In particular, equation 3.9 implies $za \circ z'$ is quasi-regular. Now the $\circ$ product of two right quasi-regular elements is another right quasi-regular element. Therefore,

$$za = za \circ 0 = za \circ (z' \circ zeae) = (za \circ z') \circ (zeae)$$

is right quasi-regular in T. We have now shown that if $z \in J(eTe)$, then za is right quasi-regular in T for any $a \in T$. Theorem 1.6 implies $J(eTe) \subseteq J(T)$.

Now $eTe \cap \mathfrak{A} = e\mathfrak{A}e$ for any ideal $\mathfrak{A}$ in T. Therefore, $J(eTe) \subseteq eTe \cap J(T) = eJ(T)e$. Hence $J(eTe) = eJ(T)e$, and the proof of Lemma 3.6 is complete. ∎

We can apply the lemma to the ring $M_{n\times n}(R)$ and get the following theorem.

Theorem 3.10 For any commutative ring R, $J(M_{n\times n}(R)) = M_{n\times n}(J(R))$.

Proof. $J(M_{n\times n}(R)$ is an ideal in $M_{n\times n}(R)$. By Theorem 3.1, $J(M_{n\times n}(R)) = M_{n\times n}(\mathfrak{A})$ for some (unique) ideal $\mathfrak{A}$ in R. We claim $\mathfrak{A} = J(R)$.

The matrix unit E_{11} is an idempotent element of $M_{n\times n}(R)$. Hence, Lemma 3.6 implies

$$E_{11}M_{n\times n}(\mathfrak{A})E_{11} = E_{11}J(M_{n\times n}(R))E_{11} = J(E_{11}M_{n\times n}(R)E_{11})$$

By 3.2d, $E_{11}M_{n\times n}(R)E_{11} = \{rE_{11} \mid r \in R\}$. Thus, the ring $E_{11}M_{n\times n}(R)E_{11}$ is isomorphic to R via the ring homomorphism sending rE_{11} to r. This map sends $E_{11}M_{n\times n}(\mathfrak{A})E_{11}$ to $\mathfrak{A}$ and $J(E_{11}M_{n\times n}(R)E_{11})$ to $J(R)$. Therefore, $\mathfrak{A} = J(R)$. ∎

There are many simple examples of Theorem 3.10. Suppose for instance $R = F[[X]]$, the ring of formal power series in X with coefficients in a field F. The Jacobson radical of R is the principal ideal generated by X. Thus, $J(R) = (X)$. Theorem 3.10 imples the Jacobson radical of $M_{n\times n}(F[[X]])$ is the set

$$J(M_{n\times n}(F[[X]])) = \{A = (a_{ij}) \in M_{n\times n}(F[[X]]) \mid X \mid a_{ij} \text{ for all } i,j = 1, \ldots, n\}$$

EXERCISES

1. A set of matrices $\{F_{ij} \mid i,j = 1, \ldots, n\}$ in $M_{n\times n}(R)$ is called a generalized set of matrix units if the following properties are satisfied:
   (a) $F_{ij} = F_{pq}$ if and only if $(i,j) = (p,q)$.
   (b) $F_{ij}F_{pq} = \delta_{jp}F_{iq}$.
   (c) $\sum_{i=1}^{n} F_{ii} = I_n$.

These conditions are to hold for all $i,j,p,q \in \{1, \ldots, n\}$. Exhibit a set of generalized matrix units other than the matrix units given in equation 2.2.

2. Let $\{E_{ij} \mid i,j = 1, \ldots, n\}$ be the matrix units of $M_{n\times n}(R)$. For each $i \neq j$ and $r \in R$, let $V_{ij}(r) = I_n + rE_{ij}$. Let $A \in M_{n\times n}(R)$. Describe $V_{ij}(r)A$ and $AV_{ij}(r)$ in terms of the rows and columns of A, respectively. Show the following identities are true:
   (a) $V_{ij}(r)V_{ij}(s) = V_{ij}(r + s)$.
   (b) $V_{ij}(r)^{-1} = V_{ij}(-r)$.

The matrices $V_{ij}(r)$ are called elementary transvections.

3. Using the same notation as in Exercise 2, let

$$P_{ij} = I_n - E_{ii} - E_{jj} + E_{ij} + E_{ji}$$

Again describe what multiplication by P_{ij} on the left and right does to an arbitrary matrix A. Show ${P_{ij}}^2 = I_n$ and compute $P_{ij}\,\mathrm{Diag}(r_1, \ldots, r_n)\,P_{ij}$ for a diagonal matrix $\mathrm{Diag}(r_1, \ldots, r_n)$. The matrix P_{ij} is called an elementary permutation matrix.

4. Using the same notation as in Exercise 2, let $S_i(r) = I_n + (r - 1)E_{ii}$ where $r \in U(R)$. Describe $S_i(r)A$ and $AS_i(r)$ in terms of the rows and columns of A, respectively. Show the following formulas are true:
   (a) $S_i(r)S_i(s) = S_i(rs)$.
   (b) $S_i(r)^{-1} = S_i(r^{-1})$.

The matrices $S_i(r)$ are called elementary dilations. Clearly, the three types of matrices $V_{ij}(r)$, P_{ij}, and $S_i(r)$ are the analogs of the classical elementary matrices used in Gaussian elimination.

5. If $\mathfrak{A}$ is a nil ideal in R, is $M_{n\times n}(\mathfrak{A})$ a nil ideal in $M_{n\times n}(R)$?
6. Examine the proof of Theorem 3.1 carefully. Is this result true if R is replaced by an arbitrary ring T? What about Theorem 3.10?
7. Prove the assertions in 3.3.
8. Recall that an ideal $\mathfrak{A}$ in a ring T is prime if, whenever $\mathfrak{B}$ and $\mathfrak{C}$ are ideals in T and $\mathfrak{B}\mathfrak{C} \subseteq \mathfrak{A}$, then $\mathfrak{B} \subseteq \mathfrak{A}$ or $\mathfrak{C} \subseteq \mathfrak{A}$. Show:
   (a) $\mathfrak{A}$ is prime if and only if $xTy \subseteq \mathfrak{A} \Rightarrow x \in \mathfrak{A}$ or $y \in \mathfrak{A}$.
   (b) $\mathfrak{A}$ is a prime ideal of R if and only if $M_{n\times n}(\mathfrak{A})$ is a prime ideal in $M_{n\times n}(R)$.
9. Recall that an ideal $\mathfrak{A}$ in a ring T is primitive if the ring $T/\mathfrak{A}$ has a faithful, irreducible representation. Show that $\mathfrak{A}$ is a primitive ideal of T if and only if $M_{n\times n}(\mathfrak{A})$ is a primitive ideal of $M_{n\times n}(T)$.
10. Suppose R satisfies the ascending chain condition. Does $M_{n\times n}(R)$ satisfy the ascending chain condition? Does $M_{n\times n}(R)$ satisfy the ascending chain condition for left ideals?

11. Let

$$\mathbb{H} = \left\{ \begin{bmatrix} z & w \\ -\overline{w} & \overline{z} \end{bmatrix} \in M_{2\times 2}(\mathbb{C}) \mid z,w \in \mathbb{C} \right\}$$

Here $\mathbb{C}$ denotes the field of complex numbers and $\bar{z}$ denotes the complex conjugate of z. Show $\mathbb{H}$ is a simple ring of dimension 4 as a vector space over $\mathbb{R}$ (the field of real numbers). The set $\mathbb{H}$ is called the quaterions.

12. Suppose M is an irreducible R-module. Show $\mathcal{E}(M) = \mathrm{Hom}_R(M,M)$ is a simple ring.
13. Let T be an arbitrary ring. Recall an element $z \in T$ is said to be left quasi-regular if $z' \circ z = 0$ for some $z' \in$ T. The element z is said to be right quasi-regular if $z \circ w = 0$ for some $w \in T$. z is said to be quasi-regular if z is both left quasi-regular and right quasi-regular. Show:
    (a) $(T, \circ, 0)$ is an associative monoid with identity 0.
    (b) z is quasi-regular if and only if $z \circ w = w \circ z = 0$ for some $w \in T$.
    (c) The quasi-regular elements in T are precisely the units in the monoid $(T, \circ, 0)$.
    (d) Any nilpotent element in T is quasi-regular.
14. Exhibit a ring T which has at least two quasi-regular elements but has no nonzero quasi-regular ideals.
15. Let $T = M_{2\times 2}(\mathbb{Q}[[X]])$. Let

$$A = \begin{bmatrix} X & X \\ X + X^2 & X^2 \end{bmatrix} \in T$$

Explain why $Y = I_2 - A$ is invertible using Theorem 3.10. Compute an inverse for Y.

16. Let $T = M_{n\times n}(\mathbb{Z}/m\mathbb{Z})$. Compute $J(T)$. Be as specific as you can.

4

The Rank of a Matrix

Let $A \in M_{m\times n}(R)$.

Definition 4.1 For each $t = 1, \ldots, r = \min\{m,n\}$, $I_t(A)$ will denote the ideal in R generated by all $t\times t$ minors of A.

Thus, to compute $I_t(A)$, calculate the determinant of each $t\times t$ submatrix of A and then find the ideal of R these determinants generate. Laplace's theorem (Theorem 2.19) implies every $(t+1)\times(t+1)$ minor of A lies in $I_t(A)$. Thus, we have the following ascending chain of ideals in R:

4.2 $$I_r(A) \subseteq I_{r-1}(A) \subseteq \cdots \subseteq I_2(A) \subseteq I_1(A) \subseteq R$$

It will be notationally convenient to extend the definition of $I_t(A)$ to all values of $t \in \mathbb{Z}$ as follows:

4.3 $$I_t(A) = \begin{cases} (0) & \text{if } t > \min\{m,n\} \\ R & \text{if } t \leq 0 \end{cases}$$

Then we have

4.4 $$(0) = I_{r+1}(A) \subseteq I_r(A) \subseteq \cdots \subseteq I_1(A) \subseteq I_0(A) = R$$

Lemma 4.5 Let $B \in M_{m\times p}(R)$ and $C \in M_{p\times n}(R)$. Then

$$I_t(BC) \subseteq I_t(B) \cap I_t(C) \quad \text{for all } t \in \mathbb{Z}$$

Proof. We leave the trivial cases to the reader and assume $1 \leq t \leq \min\{m,n\}$. We divide the proof of the lemma into three claims, which are of some interest in their own right.

Claim I. $I_t(BC) \subseteq I_t(C)$.

To see this, partition C into columns $C = (\delta_1 \mid \delta_2 \mid \cdots \mid \delta_n)$. From equation 2.11, $BC = (B\delta_1 \mid \cdots \mid B\delta_n)$. Let Δ be a $t\times t$ minor of BC, that is, a generator of the ideal $I_t(BC)$. Suppose Δ is defined by columns numbered $j_1 < j_2 < \cdots < j_t$ of BC. Since

$$I_t((\delta_{j_1} \mid \cdots \mid \delta_{j_t})) \subseteq I_t(C)$$

and

$$\Delta \in I_t((B\delta_{j_1} \mid \cdots \mid B\delta_{j_t})) = I_t(B(\delta_{j_1} \mid \cdots \mid \delta_{j_t}))$$

it suffices to show

$$I_t(B(\delta_{j_1} \mid \cdots \mid \delta_{j_t})) \subseteq I_t(\delta_{j_1} \mid \cdots \mid \delta_{j_t})).$$

In other words, in proving $\Delta \in I_t(C)$, we can assume with no loss of generality that $t = n \leq m$. Thus, $\Delta = \Delta(i_1, \ldots, i_n; 1, \ldots, n)$ for some choice of row indices $1 \leq i_1 < i_2 < \cdots < i_n \leq m$.

Suppose $B = (b_{ij}) \in M_{m\times p}(R)$. Then equation 2.14 implies

4.6 $\operatorname{Row}_i(BC) = \sum_{j=1}^{p} b_{ij} \operatorname{Row}_j(C)$ for all $i = 1, \ldots, m$.

Using the fact that the determinant is a multilinear function of its rows, we have

4.7
$$\begin{aligned}
&\Delta(i_1, \ldots, i_n; 1, \ldots, n) \\
&\quad = \det((\operatorname{Row}_{i_1}(BC); \ldots; \operatorname{Row}_{i_n}(BC)) \\
&\quad = \det((\sum_{j=1}^{p} b_{i_1 j} \operatorname{Row}_j(C); \ldots; \sum_{j=1}^{p} b_{i_n j} \operatorname{Row}_j(C))) \\
&\quad = \sum_{\alpha_1,\ldots,\alpha_n=1}^{p} c_{\alpha_1\ldots\alpha_n} \det((\operatorname{Row}_{\alpha_1}(C); \ldots; \operatorname{Row}_{\alpha_n}(C)))
\end{aligned}$$

In equation 4.7, the $c_{\alpha_1\ldots\alpha_n}$ are various constants in R coming from the expansion of the determinant. The symbols $\sum_{\alpha_1,\ldots,\alpha_n=1}^{p}$ mean the sum is taken over all

indices $\alpha_1, \ldots, \alpha_n$. For each $i = 1, \ldots, n$, the index α_i ranges from 1 to p. For all choices of $\alpha_1, \ldots, \alpha_n$, $\det((\mathrm{Row}_{\alpha_1}(C); \ldots; \mathrm{Row}_{\alpha_n}(C))) \in I_n(C)$. These determinants are all zero if $n > p$. At any rate, $\Delta(i_1, \ldots, i_n; 1, \ldots, n) \in I_t(C)$. Since Δ is an arbitrary generator of $I_t(BC)$, we conclude that $I_t(BC) \subseteq I_t(C)$.

Claim 2. $I_\alpha(A^t) = I_\alpha(A)$ for all $\alpha \in \mathbb{Z}$.
This is clear from the definitions.

Claim 3. $I_t(BC) \subseteq I_t(B)$ for all $t \in \mathbb{Z}$.
Let $\alpha \in \mathbb{Z}$. Using claims 1 and 2, we have

$$I_\alpha(BC) = I_\alpha((BC)^t) = I_\alpha(C^tB^t) \subseteq I_\alpha(B^t) = I_\alpha(B)$$

This proves claim 3. Obviously, the lemma follows from claims 1 and 3. ∎

The most important application of Lemma 4.5 is the following corollary.

Corollary 4.8 Let $A \in M_{m\times n}(R)$, $P \in \mathrm{Gl}(m,R)$, and $Q \in \mathrm{Gl}(n,R)$. Then

$$I_t(PAQ) = I_t(A) \quad \text{for all } t \in \mathbb{Z}$$

Proof. By Lemma 4.5

$$I_t(PA) \subseteq I_t(A) = I_t(P^{-1}(PA)) \subseteq I_t(PA)$$

Therefore, $I_t(PA) = I_t(A)$ for all $t \in \mathbb{Z}$. A similar proof shows $I_t(PA) = I_t(PAQ)$. ∎

We can now define the rank of a matrix $A \in M_{m\times n}(R)$. Consider the ascending sequence of ideals given in equation 4.4. Computing the annihilator of each ideal in 4.4, we get the following ascending chain of ideals.

4.9 $(0) = \mathrm{Ann}_R(R) \subseteq \mathrm{Ann}_R(I_1(A)) \subseteq \mathrm{Ann}_R(I_2(A)) \subseteq \cdots \subseteq \mathrm{Ann}_R(I_r(A)) \subseteq \mathrm{Ann}_R((0)) = R$

Notice if $\mathrm{Ann}_R(I_t(A)) \neq (0)$, then $\mathrm{Ann}_R(\mathrm{I_k}(A)) \neq (0)$ for all $k \geq t$. Thus, the following definition makes perfectly good sense.

Definition 4.10 Let $A \in M_{m\times n}(R)$. The rank of A, denoted by $\mathrm{rk}(A)$, is the following integer: $\mathrm{rk}(A) = \max\{t \mid \mathrm{Ann}_R(I_t(A)) = (0)\}$.

The are several rather obvious remarks about $\mathrm{rk}(A)$ which follow directly from the definition. We list these remarks next.

4.11 Let $A \in M_{m\times n}(R)$.
(a) $0 \leq \mathrm{rk}(A) \leq \min\{m,n\}$.
(b) $\mathrm{rk}(A) = \mathrm{rk}(A^t)$.
(c) $\mathrm{rk}(A) = \mathrm{rk}(PAQ)$ for any $P \in \mathrm{Gl}(m,R)$ and $Q \in \mathrm{Gl}(n,R)$.
(d) $\mathrm{rk}(A) = 0$ if and only if $\mathrm{Ann}_R(I_1(A)) \neq (0)$.
(e) If $m = n$, then $\mathrm{rk}(A) < n$ if and only if $\det(A) \in Z(R)$.

$I_0(A) = R$, and $\mathrm{Ann}_R(R) = 0$. Thus, $\mathrm{rk}(A) \geq 0$. On the other hand, if $t > \min\{m,n\}$, then $I_t(A) = (0)$ and $\mathrm{Ann}_R((0)) = R$. Therefore, $\mathrm{rk}(A) \leq \min\{m,n\}$. This proves 4.11a. Since $I_\alpha(A) = I_\alpha(A^t)$ for all $\alpha \in \mathbb{Z}$, 4.11b is clear. The assertion in 4.11c follows directly from Corollary 4.8. The assertions in 4.11d and 4.11e follow directly from the definition.

Let $A = (a_{ij}) \in M_{m\times n}(R)$. It follows from 4.11d that $\mathrm{rk}(A) = 0$ if and only if there exists a nonzero $x \in R$ such that $xa_{ij} = 0$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$. In particular, unlike the classical case, a matrix can have rank zero without being the zero matrix. Consider the following concrete example.

Example 4.12 Let $R = \mathbb{Z}/6\mathbb{Z} = \{0,1,2,3,4,5\}$.
(a) Suppose

$$A = \begin{bmatrix} 2 & 2 \\ 0 & 2 \end{bmatrix} \in M_{2\times 2}(R)$$

Clearly A is a nonzero matrix. Every entry in A is a zero divisor in R. $I_2(A) = 4R$, $I_1(A) = 2R$, and $\mathrm{Ann}_R(4R) = 3R \neq (0)$, $\mathrm{Ann}_R(2R) = 3R \neq (0)$. Thus, $\mathrm{rk}(A) = 0$.
(b) Let

$$B = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \in M_{2\times 2}(R)$$

Again every entry in B is a zero divisor in R. Since $\det(B) = (0)$, 4.11e implies $\mathrm{rk}(B) < 2$. Since $I_1(B) = 2R + 3R = R$, $\mathrm{Ann}_R(I_1(B)) = (0)$. Therefore, $\mathrm{rk}(B) = 1$.
(c) Suppose

$$C = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} \in M_{2\times 2}(R)$$

Then $\det(C) = 5 \in U(R)$.
Therefore, $\mathrm{rk}(C) = 2$ by 4.11e. ∎

We next discuss the relationship between the classical rank of a matrix when R is a field and Definition 4.10. Suppose F is a field and $A \in M_{m\times n}(F)$. For the time being, let us denote the classical rank of A by $\text{rank}_F(A)$. In most elementary textbooks in linear algebra, $\text{rank}_F(A)$ is defined to be the maximum number of linearly independent rows (or columns) of A. It is well known that $\text{rank}_F(A)$ is the largest integer t such that A contains a $t\times t$ submatrix whose determinant is nonzero. (See [3, Chapter 3, Thm.3.22].) Since F is a field, $\text{Ann}_F(I_t(A)) = (0)$ if and only if $I_t(A) \neq (0)$. Thus, $\text{rk}(A)$ is the largest integer t such that A contains a $t\times t$ submatrix whose determinant is nonzero. In other words, $\text{rk}(A) = \text{rank}_F(A)$. Thus, when R is a field the definition of rank given in 4.10 agrees with the classical definition of rank.

We can carry this discussion one step further. Suppose R is an integral domain with quotient field F. Let $A \in M_{m\times n}(R)$. Since $R \subseteq F$, $M_{m\times n}(R) \subseteq M_{m\times n}(F)$, and we can view A as a matrix in $M_{m\times n}(F)$. Since R is an integral domain, $\text{Ann}_R(I_t(A)) = (0)$ if and only if $I_t(A) \neq (0)$. Thus, $\text{rk}(A) = \max\{t \mid A \text{ has a nonzero } t\times t \text{ minor}\}$. Now this number $\max\{t \mid A \text{ has a nonzero } t\times t \text{ minor}\}$ is the same whether we view A as a matrix in $M_{m\times n}(R)$ or $M_{m\times n}(F)$. Thus, $\text{rk}(A)$ is just the classical rank of A when A is viewed as a matrix in $M_{m\times n}(F)$. In symbols, we can state this result as follows:

4.13 Let R be an integral domain with quotient field F. Let $A \in M_{m\times n}(R)$. Then $\text{rk}(A) = \text{rank}_F(A)$.

We finish this section with one more familiar result about ranks.

Lemma 4.14 Let $B \in M_{m\times p}(R)$ and $C \in M_{p\times n}(R)$. Then

$$\text{rk}(BC) \leq \min\{\text{rk}(B), \text{rk}(C)\}$$

Proof. We will argue $\text{rk}(BC) \leq \text{rk}(C)$. The other inequality is argued in a similar fashion. From Lemma 4.5, we have the following inclusions:

4.15
$$R \supseteq I_1(C) \supseteq I_2(C) \supseteq \cdots$$
$$R \supseteq I_1(BC) \supseteq I_2(BC) \supseteq \cdots$$

Taking the annihilators of the sequences in 4.15 produces the following ascending chains of ideals in R:

4.16
$$(0) \subseteq \text{Ann}_R(I_1(C)) \subseteq \text{Ann}_R(I_2(C)) \subseteq \cdots$$
$$(0) \subseteq \text{Ann}_R(I_1(BC)) \subseteq \text{Ann}_R(I_2(BC)) \subseteq \cdots$$

Now suppose $\mathrm{rk}(C) = q$. Then $\mathrm{Ann}_R(I_q(C)) = (0)$, but $\mathrm{Ann}_R(I_{q+k}(C)) \neq (0)$ for all $k > 0$. Since $\mathrm{Ann}_R(I_{q+k}(C)) \subseteq \mathrm{Ann}_R(I_{q+k}(BC)$, we conclude $\mathrm{Ann}_R(I_{q+k}(BC)) \neq (0)$ for all $k > 0$. Therefore, $\mathrm{rk}(BC) \leq q$. ∎

EXERCISES

1. Construct examples where the chain in 4.2 is strictly ascending.
2. Verify Lemma 4.5 in the special cases $t \leq 0$ and $t > \min\{m,n\}$.
3. Let $A \in M_{m\times n}(R)$. Let O be the zero matrix of size $m \times p$. Show $I_t(A \mid \mathrm{O}) = I_t(A)$ for all $t \in \mathbb{Z}$.
4. Prove the assertions in 4.11d and 4.11e.
5. Let $A \in M_{n\times n}(R)$. Show $\det(A) \in U(R) \Rightarrow \mathrm{rk}(A) = n$. Is the converse of this statement true?
6. Let $R = \mathbb{Z}/12\mathbb{Z}$. Compute the rank of the following matrices with entries from R:

(a) $\begin{bmatrix} 2 & 2 \\ 4 & 6 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 3 & 5 \\ 2 & 0 & 2 \\ 2 & 4 & 6 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 4 & 2 & 7 \\ 2 & 1 & 8 & 2 \\ 3 & 0 & 3 & 4 \end{bmatrix}$

7. Compute the determinant of each of the elementary matrices $V_{ij}(r)$, P_{ij}, and $S_i(r)$ given in Exercises 2, 3, and 4 in Chapter 3. Show each of these matrices has rank n.
8. Complete the proof of Lemma 4.14 by showing $\mathrm{rk}(BC) \leq \mathrm{rk}(B)$.
9. Let F be a field. Let $A_1, A_2 \in M_{m\times n}(F)$ and $B \in M_{n\times p}(F)$. Prove the following classical results about rank:
   (a) $\mathrm{rank}_F(A_1 + A_2) \leq \mathrm{rank}_F(A_1) + \mathrm{rank}_F(A_2)$.
   (b) $\mathrm{rank}_F(A_1) + \mathrm{rank}_F(B) - n \leq \mathrm{rank}_F(A_1B)$.
10. Are the classical results given in Exercise 9 true for an arbitrary commutative ring R and the function rk? Give proofs or counterexamples.
11. Let $A \in M_{m\times n}(R)$. Suppose we define $\mathrm{rk}_1(A) = \max\{t \mid I_t(A) \neq (0)\}$. Show
    (a) $0 \leq \mathrm{rk}_1(A) \leq \min\{m,n\}$.
    (b) $\mathrm{rk}_1(A) = \mathrm{rk}_1(A^t)$.
    (c) $\mathrm{rk}_1(A) = \mathrm{rk}_1(PAQ)$ for any $P \in \mathrm{Gl}(m,R)$ and $Q \in \mathrm{Gl}(n,R)$.
    (d) $\mathrm{rk}(A) \leq \mathrm{rk}_1(A)$.
    (e) $\mathrm{rk}(A) = \mathrm{rk}_1(A) = \mathrm{rank}_F(A)$ for any integral domain A with quotient field F.
12. Give an example where $\mathrm{rk}(A) < \mathrm{rk}_1(A)$ in Exercise 11.
13. Let $A \in M_{m\times n}(R)$ and $B \in R^m$. If R is a field, then the equation $AX = B$ has a solution if and only if $\mathrm{rank}_F(A \mid B) = \mathrm{rank}_F(A)$. Give a proof of this

fact. Give an example which shows that for commutative rings in general, $\mathrm{rk}(A \mid B) = \mathrm{rk}(A)$ need not imply $AX = B$ has a solution in R^n.

14. Let $A \in M_{n\times n}(R)$, and $B \in R^n$. Suppose the equation $AX = B$ has a solution $\xi \in R^n$. Show ξ is unique if and only if $\mathrm{rk}(A) = n$.
15. Formulate and prove a suitable generalization of the result in Exercise 14 when A is not square.
16. In Exercise 14, if $\mathrm{rk}(A) = n$, can we conclude $AX = B$ has a solution for all $B \in R^n$?
17. In the classical case, $\mathrm{rank}_F(A)$ is the maximum number of linearly independent columns (or rows) of A. Is $\mathrm{rk}(A)$ the maximum number of linearly independent columns (or rows) of A for an arbitrary commutative ring R?

5
Linear Equations

In this chapter, we present the basic theorems on linear systems of equations over a commutative ring R. Consider the following system of equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\quad & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots\quad a_{mn}x_n &= b_m \end{aligned} \tag{5.1}$$

The linear system in 5.1 represents m equations in the variables (i.e., unknowns) $x_1, \ldots, x_n$. The coefficients a_{ij} of these equations are elements from R, and the constants $b_1, \ldots, b_m$ are also elements from R. The equations in 5.1 can be written succinctly in matrix form as follows:

$$AX = B. \tag{5.2}$$

Here $A = (a_{ij}) \in M_{m\times n}(R), B = (b_1, \ldots, b_m)^t \in R^m$, and $X = (x_1, \ldots, x_n)^t \in (R[x_1, \ldots, x_n])^n$. Remember $x_1, \ldots, x_n$ are indeterminates over R in this notation.

The equations in 5.1 (or 5.2) have a solution (in R^n) if there exists a vector $\xi \in R^n$ such that $A\xi = B$. If $B = O$, then the system $AX = O$ is called a

homogeneous system of equations. A homogeneous system of equations $AX = O$ always has at least one solution, namely $\xi = O = (0, \ldots, 0)^t \in R^n$. We will call $\xi = O$ the trivial solution to $AX = O$. A vector $\xi \in R^n$ will be called a nontrivial solution of $AX = O$, if $\xi \neq O$, and $A\xi = O$.

Our first result in this section is a famous theorem by N. McCoy. It tells us precisely when a homogeneous system of equations $AX = O$ has a nontrivial solution.

Theorem 5.3 (N. McCoy) Let $A \in M_{m\times n}(R)$. The homogeneous system of equations $AX = O$ has a nontrivial solution if and only if $\mathrm{rk}(A) < n$.

Proof. Suppose $AX = O$ has a nontrivial solution $\xi \in R^n$. Since $\xi \neq O$, some coordinate of ξ, say $[\xi]_k$, is not zero. If $m < n$, then 4.11a implies $\mathrm{rk}(A) \leq \min\{m,n\} = m < n$. Hence, there is nothing to prove in this case. Thus, we may assume $m \geq n$.

Let $\Delta(i_1, \ldots, i_n; 1, \ldots, n)$ be an $n \times n$ minor of A. There exists a permutation matrix $P \in \mathrm{Gl}(m,R)$ such that PA has rows $i_1, \ldots, i_n$ of A as its first n rows. Thus, $\mathrm{Row}_1(PA) = \mathrm{Row}_{i_1}(A)$, $\mathrm{Row}_2(PA) = \mathrm{Row}_{i_2}(A)$, $\ldots$, and $\mathrm{Row}_n(PA) = \mathrm{Row}_{i_n}(A)$. We have the following picture of PA.

$$5.4 \qquad PA = \left[\begin{array}{ccc} a_{i_1 1} & a_{i_1 2} & \cdots \; a_{i_1 n} \\ a_{i_2 1} & a_{i_2 2} & \cdots \; a_{i_2 n} \\ \vdots & \vdots & \vdots \\ a_{i_n 1} & a_{i_n 2} & \cdots \; a_{i_n n} \\ \hline & * & \end{array}\right]$$

Set

$$D = \begin{bmatrix} a_{i_1 1} & \cdots & a_{i_1 n} \\ a_{i_2 1} & \cdots & a_{i_2 n} \\ \vdots & & \vdots \\ a_{i_n 1} & \cdots & a_{i_n n} \end{bmatrix}$$

Then $\Delta = \det(D) = \Delta(i_1, \ldots, i_n; 1, \ldots, n)$. Since $A\xi = O$, $D\xi = O$. Equation 2.20 then implies $\Delta\xi = (\Delta I_n)\xi = (\mathrm{adj}(D))D\xi = O$. In particular, $\Delta[\xi]_k = 0$. Since $\Delta = \Delta(i_1, \ldots, i_n; 1, \ldots, n)$ is an arbitrary $n \times n$ minor of A, we conclude $[\xi]_k \in \mathrm{Ann}_R(I_n(A))$. Thus, $\mathrm{Ann}_R(I_n(A)) \neq (0)$, and $\mathrm{rk}(A) < n$.

Conversely, suppose $\mathrm{rk}(A) = r < n$. If $r = m$, we can add more equations to 5.1 with zero coefficients. We get a new system of equations whose matrix form looks like the following:

$$\mathbf{5.5}\qquad \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \\ \\ \\ \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Obviously, any nonzero solution $\xi \in R^n$ of equation 5.5 is a nonzero solution to $AX = \mathrm{O}$ and vice versa. It follows from Exercise 3 of the last chapter, that

$$I_t\left(\frac{A}{\mathrm{O}}\right) = I_t(A)$$

for any $p \times n$ zero matrix O and any $t \in \mathbb{Z}$. Therefore,

$$rk(A) = rk\left(\frac{A}{\mathrm{O}}\right)$$

Thus, by replacing the equations in 5.1 with those in 5.5 if need be, we can assume $r < \min\{m,n\}$.

Since $\mathrm{rk}(A) = r$, the ideal $\mathrm{Ann}_R(I_{r+1}(A))$ is not zero. Let a be a nonzero element in $\mathrm{Ann}_R(I_{r+1}(A))$. If $r = 0$, then $a \in \mathrm{Ann}_R(I_1(A))$. In particular, $\xi = (a, \ldots, a)^t \in R^n$ is a nontrivial solution to $AX = \mathrm{O}$. Hence, we can assume $1 \leq r < \min\{m,n\}$.

Since $\mathrm{rk}(A) = r$, the ideal $\mathrm{Ann}_R(I_r(A))$ is zero. In particular, there exists an $r \times r$ minor $\Delta(i_1, \ldots, i_r; j_1, \ldots, j_r)$ of A such that $a\Delta(i_1, \ldots, i_r; j_1, \ldots, j_r) \neq 0$. We can move rows $i_1, \ldots, i_r$ and columns $j_1, \ldots, j_r$ of A to the first r rows and first r columns of a new matrix by multiplying A on the left and right with suitable permutation matrices. Thus, there exist permutation matrices $P \in \mathrm{Gl}(m,R)$ and $Q \in \mathrm{Gl}(n,R)$ such that

$$\mathbf{5.6}\quad PAQ = \left[\begin{array}{c|c} C & * \\ \hline * & * \end{array}\right] \quad \text{with } C \in M_{r\times r}(R) \text{ and } \det(C) = \Delta(i_1, \ldots, i_r; j_1, \ldots, j_r)$$

Suppose the equation $(PAQ)X = O$ has a nontrivial solution $\beta \in R^n$. Since P and Q are invertible, $\xi = Q\beta \neq O$, and $A\xi = O$. Hence $AX = O$ has a nontrivial solution. Notice $I_t(PAQ) = I_t(A)$ for all $t \in \mathbb{Z}$ by Corollary 4.8. Thus, it suffices to show $(PAQ)X = O$ has a nontrivial solution. In other words, replacing A with PAQ if need be, we can assume with no loss of generality that

$$\Delta(i_1, \ldots, i_r; j_1, \ldots, j_r) = \Delta(1, \ldots, r; 1, \ldots, r)$$

Let $\Delta = \Delta(1, \ldots, r; 1, \ldots, r)$. Then we have

$$\textbf{5.7}\quad A = \left[\begin{array}{c|c} C & * \\ \hline * & * \end{array}\right] \quad \text{with} \quad C = \begin{bmatrix} a_{11} & \cdots & a_{1r} \\ \vdots & & \vdots \\ a_{r1} & \cdots & a_{rr} \end{bmatrix}$$

and $\det(C) = \Delta$. Also, $a\Delta \neq 0$.

$$\text{Set} \quad C' = \begin{bmatrix} a_{11} & \cdots & a_{1r} & a_{1r+1} \\ \vdots & & \vdots & \vdots \\ a_{r1} & \cdots & a_{rr} & a_{rr+1} \\ a_{r+11} & \cdots & a_{r+1r} & a_{r+1r+1} \end{bmatrix} \in M_{(r+1)\times(r+1)}(R)$$

Set $d_j = \operatorname{cof}_{r+1,j}(C')$ for $j = 1, \ldots, r+1$. Thus, $d_1, \ldots, d_{r+1}$ are the cofactors of the last row of C'. By Laplace's expansion, we have

$$\textbf{5.8}\quad \sum_{j=1}^{r+1} a_{r+1j}d_j = \det(C') \in I_{r+1}(A)$$

Let $\xi = (ad_1, \ldots, ad_{r+1}, 0, \ldots, 0)^t \in R^n$. Notice that $\xi \neq O$ since $ad_{r+1} = a\Delta \neq 0$. We claim ξ is a solution to $AX = O$.

$A\xi = O$ if and only if $\sum_{j=1}^{r+1} a_{ij}(ad_j) = 0$ for all $i = 1, \ldots, m$. There are two cases to consider here. Suppose $1 \leq i \leq r$. Then

$$\sum_{j=1}^{r+1} a_{ij}(ad_j) = a\left(\sum_{j=1}^{r+1} a_{ij}d_j\right) = 0$$

by 2.19a. If $i \geq r+1$,

$$\sum_{j=1}^{r+1} a_{ij}(ad_j) = a \det \begin{bmatrix} a_{11} & \cdots & a_{1r+1} \\ \vdots & & \vdots \\ a_{r1} & \cdots & a_{rr+1} \\ a_{i1} & \cdots & a_{ir+1} \end{bmatrix} \in aI_{r+1}(A) = (0)$$

Therefore, $A\xi = O$, and the proof of Theorem 5.3 is complete. ∎

There are many interesting theorems whose proofs are direct consequences of McCoy's theorem. We list some of these results next. We begin with a generalization of a familiar result from classical linear algebra.

Corollary 5.9 Any homogeneous system of linear equations has a nontrivial solution if the number of equations is less than the number of unknowns.

Proof. Let $AX = O$ be the matrix representation of a system of linear equations in which the number of equations is less than the number of unknowns. If $A \in M_{m\times n}(R)$, then $m < n$. By 4.11a, $\text{rk}(A) \leq \min\{m,n\} = m < n$. Thus, $AX = O$ has a nontrivial solution by Theorem 5.3. ∎

McCoy's theorem can be used to analyze free R-module bases of finitely generated, free R-modules. Our first theorem in this direction leads to the definition of the rank of a free R-module.

Theorem 5.10 Let M be a finitely generated R-module. Suppose $m_1, \ldots, m_k$ and $p_1, \ldots, p_n$ are elements in M such that $\{m_1, \ldots, m_k\}$ is linearly independent over R, and $\{p_1, \ldots, p_n\}$ is a set of generators of M. Then $k \leq n$. Furthermore, if $k = n$, then $\{p_1, \ldots, p_n\}$ is a free R-module basis of M.

Proof. Since $\{p_1, \ldots, p_n\}$ is a basis of M, there exist elements $a_{ij} \in R$ such that $m_j = \sum_{i=1}^{n} a_{ij}p_i$ for $j = 1, \ldots, k$. Set $A = (a_{ij}) \in M_{n\times k}(R)$. If $k > n$, then Corollary 5.9 implies $AX = O$ has a nontrivial solution $\xi = (r_1, \ldots, r_k)^t \in R^k$. But then

$$\textbf{5.11} \quad \sum_{j=1}^{k} r_j m_j = \sum_{j=1}^{k} r_j \left(\sum_{i=1}^{n} a_{ij}p_i \right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{k} a_{ij}r_j \right) p_i = \sum_{i=1}^{n} (0)p_i = 0$$

Since $\{m_1, \ldots, m_k\}$ is linearly independent, equation 5.11 implies $r_1 = \cdots = r_k = 0$. This is clearly impossible, since at least one entry of ξ must be nonzero. Thus, $k \leq n$, and the first part of Theorem 5.10 is proved.

Now suppose $k = n$. Thus, suppose M contains n elements $m_1, \ldots, m_n$ which are linearly independent over R and M contains n elements $p_1, \ldots, p_n$ which generate M as an R-module. We will argue $p_1, \ldots, p_n$ are linearly independent

over R and, consequently, form a free R-module basis of M. Again let $m_j = \sum_{i=1}^{n} a_{ij}p_i$ for $j = 1, \ldots, n$. Set $A = (a_{ij}) \in M_{n\times n}(R)$. The computation in equation 5.11 shows the homogeneous system of equations $AX = O$ has no nontrivial solution. Therefore, Theorem 5.3 implies $\mathrm{rk}(A) = n$.

We now pass to the total quotient ring $Q(R)$ of R (see Example 13.36 with $M = R$ and $S =$ the regular elements of R). Recall $Q(R) = \{x/y \mid x,y \in R$, and y regular$\}$. The ring R is identified with the subring $\{x/1 \mid x \in R\} \subseteq Q(R)$. Then $A \in M_{n\times n}(R) \subseteq M_{n\times n}(Q(R))$. Since $\mathrm{rk}(A) = n$, 4.11e implies $\det(A)$ is a regular element in R. In particular, $\det(A) \in U(Q(R))$. It follows from Corollary 2.21 that A is invertible in $M_{n\times n}(Q(R))$. Let $B = (b_{ij}) \in M_{n\times n}(Q(R))$ be the inverse of A.

Suppose $y_1p_1 + \cdots + y_np_n = 0$ for some $(y_1, \ldots, y_n)^t = \xi \in R^n$. Set $c_j = \sum_{k=1}^{n} y_kb_{jk}$ for $j = 1, \ldots, n$. Then $(c_1, \ldots, c_n)^t = B\xi \in (Q(R))^n$. Since $AB = I_n$ in $M_{n\times n}(Q(R))$, we have

$$\textbf{5.12}\quad \sum_{j=1}^{n} c_jm_j = \sum_{j=1}^{n} c_j\left(\sum_{i=1}^{n} a_{ij}p_i\right) = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}c_j\right)p_i$$

$$= \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}\left(\sum_{k=1}^{n} y_kb_{jk}\right)\right)p_i$$

$$= \sum_{i=1}^{n}\left(\sum_{k=1}^{n} y_k\left(\sum_{j=1}^{n} a_{ij}b_{jk}\right)\right)p_i = \sum_{i=1}^{n}\left(\sum_{k=1}^{n} y_k\delta_{ik}\right)p_i$$

$$= \sum_{i=1}^{n} y_ip_i = 0$$

Any finite number of elements in $Q(R)$ have a common denominator, so there is a regular element $x \in R$ such that $xc_1, \ldots, xc_n \in R$. From equation 5.12, we have $\sum_{j=1}^{n}(xc_j)m_j = 0$. Since $m_1, \ldots, m_n$ are linearly independent over R, we conclude $xc_1 = \cdots = xc_n = 0$. Since x is a regular element of R, x is a unit in $Q(R)$. Therefore, $c_1 = \cdots = c_n = 0$. Since B is invertible, $\xi = O$. Thus, $y_1 = \cdots = y_n = 0$. We conclude $p_1, \ldots, p_n$ are linearly independent over R. This completes the proof of Theorem 5.10. ∎

Another result closely related to Theorem 5.10 is the following statement.

Corollary 5.13 If $R^n \cong R^m$ as R-modules, then $n = m$.

Proof. Suppose η is an R-module isomorphism from R^n onto R^m. Let $\varepsilon = \{\varepsilon_1, \ldots, \varepsilon_n\}$ denote the canonical basis of R^n. Thus, $\varepsilon_1 = (1,0, \ldots, 0)^t$, $\varepsilon_2 = (0,1,0, \ldots, 0)^t, \ldots, \varepsilon_n = (0, \ldots, 0,1)^t$ in R^n. Since η is an isomorphism, $\eta(\varepsilon_1), \ldots, \eta(\varepsilon_n)$ are linearly independent in R^m. The canonical basis of R^m

contains m vectors. Thus, Theorem 5.10 implies $n \leq m$. Reversing the roles of R^n and R^m in this proof gives us $m \leq n$. Therefore, $m = n$. ∎

Notice that Theorem 5.10 (or Corollary 5.13) implies any two free R-module bases of a finitely generated, free R-module M have the same cardinality. This is the analog of the classical result which says any two bases of a finite-dimensional vector space over a field must have the same cardinality, namely the dimension of V. The common cardinality of any free R-module basis of a finitely generated, free R-module M will be called the rank of M and written rank (M). For example, $\text{rank}(R^n) = n$ and $\text{rank}(M_{m\times n}(R)) = mn$. If R is a field and V is a finite-dimensional vector space over R, then $\text{rank}(V)$ is just the usual vector space dimension of V over R.

There is one more corollary to Theorem 5.10 which we want to mention here.

Corollary 5.14 Let $P,Q \in M_{m\times n}(R)$ such that $\text{CS}(P) = \text{CS}(Q)$ in R^m. If the columns of P are linearly independent in R^m, then there exists an $S \in \text{Gl}(n,R)$ such that $P = QS$.

Proof. Let $P = (\delta_1|\cdots|\delta_n)$ and $Q = (\lambda_1|\cdots|\lambda_n)$ be column partitions of P and Q, respectively. Then $R\delta_1 + \cdots + R\delta_n = \text{CS}(P) = \text{CS}(Q) = R\lambda_1 + \cdots + R\lambda_n$. Since $\delta_1, \ldots, \delta_n$ are linearly independent over R, $\{\delta_1, \ldots, \delta_n\}$ is a free R-module basis of $\text{CS}(P)$. It follows from Theorem 5.10 that $\{\lambda_1, \ldots, \lambda_n\}$ is also a free R-module basis of $\text{CS}(P)$. Let $\delta_i = \sum_{j=1}^n v_{ij}\lambda_j$ for $i = 1, \ldots, n$ and $\lambda_i = \sum_{j=1}^n t_{ij}\delta_j$ for $i = 1, \ldots, n$. Here the v_{ij} and t_{ij} are scalars from R. Set $V = (v_{ij})$ and $T = (t_{ij})$. Then $V,T \in M_{n\times n}(R)$.

Since

$$\delta_i = \sum_{j=1}^n v_{ij}\lambda_j = \sum_{j=1}^n v_{ij}\left(\sum_{k=1}^n t_{jk}\delta_k\right) = \sum_{k=1}^n\left(\sum_{j=1}^n v_{ij}t_{jk}\right)\delta_k$$

we conclude $\sum_{j=1}^n v_{ij}t_{jk} = \delta_{ik}$. Thus, $VT = I_n$. In particular, $V \in \text{Gl}(n,R)$ and $TV = I_n$. Set $S = V^t$. Then $S \in \text{Gl}(n,R)$. Using equations 2.10 and 2.11, we have

$$\begin{aligned} QS &= (Q\,\text{Col}_1(S) \mid Q\,\text{Col}_2(S) \mid \cdots \mid Q\,\text{Col}_n(S)) \\ &= \left(\sum_{j=1}^n [S]_{j1}\lambda_j \mid \cdots \mid \sum_{j=1}^n [S]_{jn}\lambda_j\right) \\ &= \left(\sum_{j=1}^n v_{1j}\lambda_j \mid \cdots \mid \sum_{j=1}^n v_{nj}\lambda_j\right) = (\delta_1 \mid \cdots \mid \delta_n) = P \end{aligned}$$

∎

Corollary 5.14 says something interesting even for 1×1 matrices, i.e., scalars in R. Suppose $a,b \in R$ and a is regular. Let $P = (a)$ and $Q = (b)$. Then the

columns of P are linearly independent in R. Suppose $CS(P) = CS(Q)$. This just says $Ra = Rb$. Corollary 5.14 implies there exists a unit $u \in U(R)$ such that $a = bu$. Thus, if $Ra = Rb$ and a is regular, then a and b are associates. We note that this result is not necessarily true if a is not regular. In other words, the hypothesis "the columns of P are linearly independent in R^m" cannot be omitted from Corollary 5.14 even in the 1×1 case. Consider the following clever example, which is due to I. Kaplansky.

Example 5.15 Let $F = \mathbb{Z}/5\mathbb{Z} = \{\bar{0},\bar{1},\bar{2},\bar{3},\bar{4}\}$. Let $n \mapsto \bar{n}$ denote the natural mapping of $\mathbb{Z}$ onto F. Set $R = \{(n,f(X)) \in \mathbb{Z} \times F[X] \mid f(0) = \bar{n} \text{ in } F\}$. It is easy to check that R is a commutative ring with addition and multiplication defined componentwise.

Let $a = (0,X)$, $b = (0,\bar{2}X)$, $z = (2,\bar{2})$, and $w = (3,\bar{3})$. Then a, b, z, and w are elements in R with $az = b$ and $bw = a$. Therefore, $Ra = Rb$. Suppose $a \sim b$. Then $(0,X) = (n,f(X))(0,\bar{2}X)$ for some unit $(n,f(X))$ in R. Now $\bar{2}Xf(X) = X$ implies $f(X) = \bar{3}$. Therefore, $(n,f(X)) = (n,\bar{3})$. But $(n,\bar{3}) \in R$ implies $n \equiv 3 \bmod 5$. In particular, $n \neq 1$ or -1. But then $(n,\bar{3})$ cannot be a unit in R; that is, there is no $(p,q(X)) \in R$ such that $(n,\bar{3})(p,q(X)) = (np,\bar{3}q(X) = (1,\bar{1})$. We conclude that a and b are not associates in R.

Set $P = (a)$ and $Q = (b)$. Then $P,Q \in M_{1\times 1}(R)$ with $CS(P) = Ra = Rb = CS(Q)$. Since $az = b$ and $wb = a$, $(wz - 1)a = 0$. Also, $wz - 1 = (3,\bar{3})(2,\bar{2}) - (1,\bar{1}) = (5,\bar{0}) \neq (0,\bar{0})$. In particular, the columns of P are not linearly independent in R^1. If $P = QS$ for some invertible matrix $S \in M_{1\times 1}(R) = R$, then $a \sim b$, which is not the case. Therefore, $P \neq QS$ for any invertible matrix S. ∎

We can use Corollary 5.14 to give the following nice characterization of invertible matrices in $M_{n\times n}(R)$.

Corollary 5.16 Let $P = (\delta_1|\cdots|\delta_n) \in M_{n\times n}(R)$. P is invertible if and only if $\{\delta_1, \ldots, \delta_n\}$ is a free R-module basis of R^n.

Proof. Suppose $\{\delta_1, \ldots, \delta_n\}$ is a free R-module basis of R^n. The columns of the identity matrix I_n are the canonical basis of R^n. Therefore, Corollary 5.14 implies $P = I_nS$ for some $S \in Gl(n,R)$. In particular, $P \in Gl(n,R)$.

Conversely, suppose $P = (\delta_1|\cdots|\delta_n) \in Gl(n,R)$. For any $\lambda \in R^n$, the equation $PX = \lambda$ has the unique solution $P^{-1}\lambda = (y_1, \ldots, y_n)^t$. Thus, $y_1\delta_1 + \cdots + y_n\delta_n = \lambda$. In particular, $\{\delta_1, \ldots, \delta_n\}$ is an R-module basis of R^n. Suppose $z_1\delta_1 + \cdots + z_n\delta_n = O$ in R^n. Let $\xi = (z_1, \ldots, z_n)^t \in R^n$. Then $P\xi = O$ and $\xi = P^{-1}(P\xi) = P^{-1}O = O$. Hence, $z_1 = z_2 = \cdots = z_n = 0$ and $\delta_1, \ldots, \delta_n$ are linearly independent over R. Therefore, $\{\delta_1, \ldots, \delta_n\}$ is a free R-module basis of R^n. ∎

In the rest of this section, we will discuss the nonhomogeneous equation $AX = B$ where B is not necessarily zero. We still have Cramer's rule, which is valid over any commutative ring R.

Theorem 5.17 (Cramer's rule) Let $A \in M_{n\times n}(R)$ with $\det(A) \in U(R)$. Then for any $B = (b_1, \ldots, b_n)^t \in R^n$, the equation $AX = B$ has the unique solution $\xi = (y_1, \ldots, y_n)^t$ where

$$y_j = (\det(A))^{-1} \det \begin{bmatrix} a_{11} & \cdots & a_{1j-1} & b_1 & a_{1j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj-1} & b_n & a_{nj+1} & \cdots & a_{nn} \end{bmatrix}$$

for all $j = 1, \ldots, n$.

Proof. Let $\xi = (y_1, \ldots, y_n)^t$ with the y_j defined as above. Using Laplace's expansion, we have

$$\textbf{5.18} \quad \det(A)y_j = \det \begin{bmatrix} a_{11} & \cdots & a_{1j-1} & b_1 & a_{1j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & & a_{nj-1} & b_n & a_{nj+1} & \cdots & a_{nn} \end{bmatrix}$$

$$= \sum_{i=1}^{n} b_i \operatorname{cof}_{ij}(A)$$

Therefore,

$$\textbf{5.19} \quad \det(A) \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \operatorname{cof}_{i1}(A)b_i \\ \\ \sum_{i=1}^{n} \operatorname{cof}_{in}(A)b_i \end{bmatrix} = \operatorname{adj}(A)B$$

Since $\det(A)I_n = \operatorname{adj}(A)A$, equation 5.19 implies

$$\textbf{5.20} \quad \operatorname{adj}(A)[A\xi] = \operatorname{adj}(A)B.$$

Since adj(A) is an invertible matrix (with inverse $(\det(A))^{-1}A$), equation 5.20 implies $A\xi = B$.

Suppose ξ' is another solution to $AX = B$. Then $A\xi' = A\xi = B$ implies $A(\xi - \xi') = O$. Since A is invertible, $\xi - \xi' = O$. Thus, ξ is the unique solution to $AX = B$. ∎

Thus, if A is invertible, $AX = B$ has a unique solution for any $B \in R^n$. Our next theorem describes necessary conditions for $AX = B$ to have a solution for any $A \in M_{m\times n}(R)$ and any $B \in R^m$.

Theorem 5.21 Let $A \in M_{m\times n}(R)$. Suppose the system of equations $AX = B$ has a solution. Then $I_t(A|B) = I_t(A)$ for all $t \in \mathbb{Z}$.

Proof. The proof of this theorem is a simple consequence of the fact that the determinant is an (alternating) multilinear function of the columns of a matrix. Before presenting the details of this argument, we take this opportunity to discuss certain standard reductions which are often employed when studying the equation $AX = B$.

If $m > n$, then we can add new variables, say $x_{n+1}, \ldots, x_m$, with zero coefficients to the equations in 5.1. The new system $A'X' = B$ has the following form.

5.22
$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & 0 & \cdots & 0 \end{bmatrix}}_{A'} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}}_{X'} = \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}}_{B}$$

$\xi = (y_1, \ldots, y_n)' \in R^n$ is a solution to $AX = B$ if and only if $\xi' = (y_1, \ldots, y_n, 0, \ldots, 0)' \in R^m$ is a solution to $A'X' = B$. Also, Exercise 3 of the last chapter implies $I_t(A) = I_t(A')$ and $I_t(A \mid B) = I_t(A' \mid B)$ for all $t \in \mathbb{Z}$. Thus, if $I_t(A' \mid B) = I_t(A')$ for all $t \in \mathbb{Z}$ then $I_t(A \mid B) = I_t(A)$ for all $t \in \mathbb{Z}$. The point of all this is that by passing to $A'X' = B$ if need be, we can assume with no loss of generality that $m \leq n$.

Since both A and $(A \mid B)$ have m rows, $I_t(A) = I_t(A \mid B) = (0)$ if $t > \min\{m,n\} = m$. Hence, we can assume, $1 \leq t \leq m = \min\{m,n\}$. Notice that $I_t(A) \subseteq I_t(A \mid B)$ for any t. This follows directly from the definitions. Hence, we need only show $I_t(A \mid B) \subseteq I_t(A)$ for $1 \leq t \leq m = \min\{m,n\}$. Obviously, any $t \times t$ minor of $(A \mid B)$ which does not involve B is contained in $I_t(A)$.

Thus, to show $I_t(A \mid B) \subseteq I_t(A)$ we need only consider those $t \times t$ minors of $(A \mid B)$ which involve the last column B of $(A \mid B)$. Such a minor has the form $\Delta(i_1, \ldots, i_t; j_1, \ldots, j_{t-1}, n+1)$. Here $1 \leq i_1 < \cdots < i_t \leq m$ and $1 \leq j_1 < \cdots < j_{t-1} \leq n$. We want to show

$$\Delta(i_1, \ldots, i_t; j_1, \ldots, j_{t-1}, n+1) \in I_t(A)$$

Fix indices $1 \leq i_1 < \cdots < i_t \leq m$ and $1 \leq j_1 < \cdots < j_{t-1} \leq n$, and consider the minor $\Delta(i_1, \ldots, i_t; j_1, \ldots, j_{t-1}, n+1) \in I_t(A \mid B)$. Let $\xi = (x_1, \ldots, x_n)^t \in R^n$ be a solution to $AX = B$. From equation 2.10 we have $x_1 \operatorname{Col}_1(A) + \cdots + x_n \operatorname{Col}_n(A) = B$ in R^m. Thus,

5.23 $\Delta(i_1, \ldots, i_t; j_1, \ldots, j_{t-1}, n+1)$

$$= \det \left[\begin{array}{ccc|c} a_{i_1 j_1}, & \ldots, & a_{i_1 j_{t-1}} & x_1 a_{i_1 1} + \cdots + x_n a_{i_1 n} \\ a_{i_2 j_1}, & \ldots, & a_{i_2 j_{t-1}} & x_1 a_{i_2 1} + \cdots + x_n a_{i_2 n} \\ \vdots & & \vdots & \vdots \\ a_{i_t j_1}, & \ldots, & a_{i_t j_{t-1}} & x_1 a_{i_t 1} + \cdots + x_n a_{i_t n} \end{array}\right]$$

$$= \sum_{k=1}^{n} x_k \det \begin{bmatrix} a_{i_1 j_1}, & \ldots, & a_{i_1 j_{t-1}}, & a_{i_1 k} \\ a_{i_2 j_1}, & \ldots, & a_{i_2 j_{t-1}}, & a_{i_2 k} \\ \vdots & & \vdots & \vdots \\ a_{i_t j_1}, & \ldots, & a_{i_t j_{t-1}}, & a_{i_t k} \end{bmatrix}$$

$$\in I_t(A) \qquad \blacksquare$$

A couple of comments concerning Theorem 5.21 are in order here. If $I_t((A \mid B)) = I_t(A)$ for all $t \in \mathbb{Z}$, then $\operatorname{rk}(A \mid B) = \operatorname{rk}(A)$. If R is a field, we could then conclude $B \in \operatorname{CS}(A)$, and consequently $AX = B$ has a solution. Thus, the converse of Theorem 5.21 is true if R is a field. In general, $I_t((A \mid B)) = I_t(A)$ for all $t \in \mathbb{Z}$ is not sufficient to guarantee $AX = B$ has a solution. Consider the following example from [4].

Example 5.24 Let F be a field, and set $R = F[X,Y]/(X^2 - Y^3)$. Here X and Y are indeterminates over F. Let x and y denote the images of X and Y, respec-

tively, in R. Since $X^2 - Y^3$ is irreducible in $F[X,Y]$, $R = F[x,y]$ is an integral domain in which $x^2 = y^3$.

Let

$$A = \begin{bmatrix} x & y & 0 \\ 0 & 0 & y \end{bmatrix} \in M_{2\times 3}(R) \quad \text{and} \quad B = \begin{bmatrix} 0 \\ x \end{bmatrix} \in R^2$$

Then $I_1(A) = (x,y) = I_1((A \mid B))$ and $I_2(A) = (xy,y^2)$. On the other hand, $I_2((A \mid B)) = (xy,x^2,y^2) = (xy,y^2) = I_2(A)$ since $x^2 = y^3 \in (xy,y^2)$. Therefore, $I_t(A) = I_t((A \mid B))$ for all $t \in \mathbb{Z}$.

If

$$\begin{bmatrix} x & y & 0 \\ 0 & 0 & y \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ x \end{bmatrix}$$

has a solution $\xi = (y_1,y_2,y_3)^t$ in R^3, then $x \in Ry$. Pulling this relation back to $F[X,Y]$, we have $X - KY = L(X^2 - Y^3)$ for some $K,L \in F[X,Y]$. This is clearly impossible. Hence, $AX = B$ has no solution in R^3. ∎

In our next theorem, which is due to Camion, Levy, and Mann, we give sufficient conditions for $AX = B$ to have a solution. In particular, we can assume $B \neq O$. As pointed out in the proof of Theorem 5.21, we can always assume $m \leq n$ when considering solutions to $AX = B$. In this case, we will let $I_m(A \mid B)^*$ denote the ideal in R generated by all $m \times m$ minors of $(A \mid B)$ which involve column B of $(A \mid B)$. In symbols, $I_m(A \mid B)^*$ is the ideal in R generated by the set

$$\{\Delta(1, \ldots, m; j_1, \ldots, j_{m-1}, n+1 \mid 1 \leq j_1 < \cdots < j_{m-1} \leq n\}$$

We then have the following result.

Theorem 5.25 Let $A \in M_{m\times n}(R)$ with $m \leq n$ and $\mathrm{rk}(A) = m$. Let $B \in R^m$. Suppose there exist an ideal $\mathfrak{A}$ in R and a regular element $z \in R$ such that $\mathfrak{A}I_m(A \mid B)^* \subseteq Rz \subseteq \mathfrak{A}I_m(A)$. Then the equation $AX = B$ has a solution.

Proof. Since $\mathrm{rk}(A) = m$, $\mathrm{Ann}_R(I_m(A)) = (0)$. In particular, $I_m(A) \neq (0)$. Hence, there exists at least one $m \times m$ minor of A which is nonzero. Suppose $\Delta = \Delta(1, \ldots, m; j_1, \ldots, j_m)$ is a nonzero $m \times m$ minor of A. As usual, we assume $1 \leq j_1 < \cdots j_m \leq n$. Consider the following $m \times m$ submatrix of A.

$$\textbf{5.26}\quad \bar{A} = \begin{bmatrix} a_{1j_1} & \cdots & a_{1j_m} \\ \vdots & & \vdots \\ a_{mj_1} & \cdots & a_{mj_m} \end{bmatrix}$$

Then $\det(\bar{A}) = \Delta \neq 0$, and

$$\textbf{5.27}\quad \Delta \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \det(\bar{A}) \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \bar{A}\ \mathrm{adj}(\bar{A}) \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

We also have

$$\textbf{5.28}\quad \mathrm{adj}(\bar{A}) \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{m} b_j \operatorname{cof}_{j1}(\bar{A}) \\ \\ \sum_{j=1}^{m} b_j \operatorname{cof}_{jm}(\bar{A}) \end{bmatrix}$$

Set

$$\begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{m} b_j \operatorname{cof}_{j1}(\bar{A}) \\ \\ \sum_{j=1}^{m} b_j \operatorname{cof}_{jm}(\bar{A}) \end{bmatrix} \in R^m$$

Then

$$c_i = \sum_{j=1}^{m} b_j \operatorname{cof}_{ji}(\overline{A}) = \det \begin{bmatrix} a_{1j_1} & \cdots & b_1 & \cdots & a_{1j_m} \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ a_{mj_1} & \cdots & b_m & \cdots & a_{mj_m} \end{bmatrix} \in I_m(A \mid B)^*$$
$$\qquad\qquad\qquad\qquad\qquad\qquad i$$

for all $i = 1, \ldots, m$. From equations 5.27 and 5.28,

$$\Delta \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{bmatrix} = \begin{bmatrix} a_{1j_1} & \cdots & a_{1j_m} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{mj_1} & \cdots & a_{mj_m} \end{bmatrix} \begin{bmatrix} c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_m \end{bmatrix}$$

Therefore,

5.29 $\Delta b_i = \sum_{u=1}^{m} a_{ij_u} c_u$ for all $i = 1, \ldots, m$.

Now define $y_1, \ldots, y_n$ by the following formulas:

5.30 $y_v = \begin{cases} 0 & \text{if } v \in \{1, \ldots, n\} - \{j_1, \ldots, j_m\} \\ c_i & \text{if } v = j_i \text{ for } i = 1, \ldots, m \end{cases}$

Notice that $y_v \in I_m(A \mid B)^*$ for all $v = 1, \ldots, n$. Also,

$$\sum_{v=1}^{n} a_{iv} y_v = a_{ij_1} y_{j_1} + \cdots + a_{ij_m} y_{j_m} = a_{ij_1} c_1 + \cdots + a_{ij_m} c_m$$
$$= \Delta b_i \quad \text{for all i} = 1, \ldots, \text{m}$$

We have now shown

5.31 $\Delta b_i = \sum_{v=1}^{n} a_{iv} y_v$ with $y_1, \ldots, y_n \in I_m(A \mid B)^*$

Equation 5.31 holds for each $i = 1, \ldots, m$.

The computation in equation 5.31 can be done for each nonzero $m \times m$ minor of A. Suppose we list the nonzero $m \times m$ minors of A as $\Delta_1, \ldots, \Delta_p$. Then for every $k = 1, \ldots, p$, there exist scalars $\{y_{kv} \in R \mid v = 1, \ldots, n\} \subseteq I_m(A \mid B)^*$ such that

5.32 $\Delta_k b_i = \sum_{v=1}^{n} a_{iv} y_{kv}$ for all $i = 1, \ldots, m$

By hypothesis, $\mathfrak{A}I_m(A \mid B)^* \subseteq Rz \subseteq \mathfrak{A}I_m(A)$. Since $\Delta_1, \ldots, \Delta_p$ generate the ideal $I_m(A)$, we have $z = \sum_{k=1}^{p} q_k\Delta_k$ for some $q_1, \ldots, q_p \in \mathfrak{A}$. Equation 5.32 implies

$$\textbf{5.33} \quad \sum_{k=1}^{p}\sum_{v=1}^{n} a_{iv}q_k y_{kv} = \sum_{k=1}^{p} q_k\left(\sum_{v=1}^{n} a_{iv}y_{kv}\right) = \sum_{k=1}^{p} q_k\Delta_k b_i = zb_i \quad \text{for all } i = 1, \ldots, m$$

Therefore,

$$\textbf{5.34} \quad \sum_{v=1}^{n} a_{iv}\left(\sum_{k=1}^{p} q_k y_{kv}\right) = zb_i \quad \text{for all } i = 1, \ldots, m$$

For each $v = 1, \ldots, n$, $\sum_{k=1}^{p} q_k y_{kv} \in \mathfrak{A}I_m(A \mid B)^* \subseteq Rz$. Hence, $\sum_{k=1}^{p} q_k y_{kv} = r_v z$ for some $r_v \in R$. Equation 5.34 then implies $z(\sum_{v=1}^{n} a_{iv}r_v) = zb_i$ for all $i = 1, \ldots, m$. Since z is a regular element of R, $\sum_{v=1}^{n} a_{iv}r_v = b_i$ for all $i = 1, \ldots, m$. Thus, $\xi = (r_1, \ldots, r_n)^t \in R^n$ is a solution to $AX = B$. This completes the proof of Theorem 5.25. ∎

One immediate application of Theorem 5.25 is the following special case. Suppose $I_m(A) = R$. Then $\mathrm{rk}(A) = m$. For any $B \in R^m$, $RI_m(A \mid B)^* \subseteq R1 \subseteq RI_m(A)$. Since 1 is a regular element of R, Theorem 5.25 implies $AX = B$ has a solution. Thus, we have proved the following corollary.

Corollary 5.35 Let $A \in M_{m\times n}(R)$ with $I_m(A) = R$. Then for any $B \in R^m$, the system of equations $AX = B$ has a solution. ∎

The theorems in this section can easily be translated into assertions about linear maps between free R-modules. Suppose $A \in M_{m\times n}(R)$. Then A induces an R-module homomorphism $\mu_A : R^n \mapsto R^m$ given by $\mu_A(\xi) = A\xi$. Theorem 5.3 and Corollary 5.35 imply the following result.

Theorem 5.36 Let $A \in M_{m\times n}(R)$. Let $\mu_A : R^n \mapsto R^m$ be the R-module homomorphism given by $\mu_A(\xi) = A\xi$.

(a) Suppose $n \geq m$. Then μ_A is surjective if and only if $I_m(A) = R$.

(b) Suppose $n \leq m$. Then μ_A is injective if and only if $\mathrm{Ann}_R(I_n(A)) = (0)$.

Before proving Theorem 5.36, let us say a few words about the hypotheses on m and n in that theorem. If $n < m$, then the map μ_A cannot be surjective by

Theorem 5.10. If $n > m$, then μ_A cannot be injective for the same reasons. Thus, the hypotheses on m and n in Theorem 5.36 are natural for deciding when μ_A is surjective or when μ_A is injective.

Proof of Theorem 5.36

(a) Suppose $n \geq m$. If $I_m(A) = R$, then Corollary 5.35 implies $AX = B$ has a solution for any $B \in R^m$. This is precisely the statement μ_A is surjective. Conversely, suppose μ_A is surjective. Let $\mathbf{\varepsilon} = \{\varepsilon_1, \ldots, \varepsilon_m\}$ denote the canonical basis of R^m. Then $AX = \varepsilon_1$ has a solution since μ_A is surjective. In particular, Theorem 5.21 implies $I_m(A) = I_m((A \mid \varepsilon_1))$. The equation $AX = \varepsilon_2$ has a solution $\xi \in R^n$ since μ_A is surjective. Then $(A \mid \varepsilon_1)Y = \varepsilon_2$ where

$$Y = \left(\frac{\xi}{0}\right)$$

In particular, the equation $(A \mid \varepsilon_1)X = \varepsilon_2$ has a solution. Again by Theorem 5.21, $I_m((A \mid \varepsilon_1)) = I_m((A \mid \varepsilon_1 \mid \varepsilon_2))$. Continuing in this fashion, we see $I_m(A) = I_m((A|\varepsilon_1 \mid \cdots \mid \varepsilon_m)) = R$.

(b) The map μ_A fails to be injective if and only if $AX = O$ has a nontrivial solution $\xi \in R^n$. By Theorem 5.3, $AX = O$ has a nontrivial solution if and only if $\text{rk}(A) < n$. This is in turn equivalent to $\text{Ann}_R(I_n(A)) \neq (0)$. This proves (b). ∎

EXERCISES

1. Show that Exercise 3 of Chapter 4 implies

$$I_t\left(\frac{A}{O}\right) = I_t(A)$$

for any zero matrix O of the appropriate size and any $t \in \mathbb{Z}$.

2. Show that any free R-module basis of a finitely generated, free R-module M is finite. Show that any two free R-module bases of M have the same cardinality.
3. Let M be a free R-module (not necessarily finitely generated). Show that any two free R-module bases of M have the same cardinality.
4. Let M and N be free R-modules of rank n and m respectively. Show $\text{Hom}_R(M,N) \cong M_{m\times n}(R)$.
5. Suppose N and M are finitely generated, free R-modules. If $N \subseteq M$, show $\text{rank}(N) \leq \text{rank}(M)$. Give an example where $N < M$ and $\text{rank}(N) = \text{rank}(M)$.
6. Check that the ring defined in Example 5.15 is a commutative ring with addition and multiplication defined componentwise. Is R a Noetherian ring? Compute $J(R)$.

7. Suppose R is a commutative ring in which $Z(R) \subsetneq J(R)$. Give a couple of different examples of such a ring. For any such ring, prove the following assertion: $x \sim y$ if and only if $Rx = Ry$.
8. Use Exercise 7 to find a ring R which is not an integral domain and has the property that $x \sim y$ if and only if $Rx = Ry$.
9. Let $R = \mathbb{Z}/4\mathbb{Z}$ and set

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 0 & 2 \end{bmatrix} \in M_{3\times 3}(R)$$

Find all solutions to $AX = B$ where $B = (1,2,3)^t \in R^3$.

10. Let

$$A = \begin{bmatrix} x & y & 0 \\ 0 & 0 & y \end{bmatrix}$$

as in Example 5.24.
Consider the homomorphism $\mu_A : R^3 \mapsto R^2$.
(a) Compute $\mathrm{Ker}(\mu_A)$.
(b) Compute $\mathrm{Im}(\mu_A)$.
(c) Find all solutions to $AX = B$ when $B = (2x^2, x^2)^t \in R^2$.

11. Let $A \in M_{n\times n}(R)$. An element $d \in R$ is called an eigenvalue of A if $\mathrm{Ker}(dI_n - A) \neq (0)$. Let $R = \mathbb{Z}/6\mathbb{Z}$. Compute all eigenvalues of

$$A = \begin{bmatrix} 1 & 0 \\ 4 & 2 \end{bmatrix} \in M_{3\times 3}(R)$$

Do the same for

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{bmatrix} \in M_{3\times 3}(R)$$

12. Let $A \in M_{m\times n}(R)$. Explain carefully why $\mu_A : R^n \mapsto R^m$ is not surjective if $n < m$. Why is μ_A not injective if $n > m$?
13. In Exercise 12, show μ_A is injective if and only if the columns of A are linearly independent in R^m.
14. In Exercise 12, show μ_A is surjective if and only if for each maximal ideal $\mathfrak{P} \subset R$, the induced map $\overline{\mu}_A : (R/\mathfrak{P})^n \mapsto (R/\mathfrak{P})^m$ is surjective.
15. Is Exercise 14 still true if the word surjective is replaced by injective?

6

Minimal Primes and the Radical of an Ideal

In this chapter, we will discuss various properties of prime ideals in commutative rings. This material will be used in the next chapter when discussing the relationships between the characteristic polynomial of a matrix and its order ideal. We begin with the definition of a prime ideal.

Let R be a commutative ring.

Definition 6.1 An ideal $\mathfrak{P}$ of R is called a prime ideal if $\mathfrak{P} \neq R$ and whenever $xy \in \mathfrak{P}$, then $x \in \mathfrak{P}$ or $y \in \mathfrak{P}$.

Notice that the improper ideal R is not a prime ideal. Clearly, a proper ideal $\mathfrak{P}$ is a prime ideal if and only if the quotient ring $R/\mathfrak{P}$ is an integral domain. In particular, (0) is a prime ideal if and only if R is an integral domain. If $\mathfrak{P}$ is a prime ideal of R and $\mathfrak{A}$ and $\mathfrak{B}$ are ideals of R for which $\mathfrak{AB} \subseteq \mathfrak{P}$, then either $\mathfrak{A} \subseteq \mathfrak{P}$ or $\mathfrak{B} \subseteq \mathfrak{P}$. This assertion follows directly from the definition.

There is an intimate relationship between prime ideals of R and multiplicatively closed subsets of R.

Definition 6.2 A subset $S \subseteq R$ is said to be multiplicatively closed if $1 \in S$ and $xy \in S$ whenever $x,y \in S$.

In this textbook, a multiplicately closed subset of R must contain 1. For example, (0) is closed under multiplication, but (0) is not a multiplicatively closed subset of R. If Γ is any subset of R, we will let $\Gamma^c = \{x \in R \mid x \notin \Gamma\}$ denote the complement of Γ.

Let $\mathfrak{P}$ be a proper ideal of R. It follows from Definition 6.1 that $\mathfrak{P}$ is a prime ideal of R if and only if $\mathfrak{P}^c$ is a multiplicatively closed subset of R. Thus, complements of prime ideals are multiplicatively closed subsets of R. There are other important examples as well. The regular elements of R, that is, $Z(R)^c$, clearly form a multiplicatively closed subset of R. If $x \in R^*$, then $S = \{x^i \mid i \geq 0\}$ is a multiplicatively closed subset of R. If $\mathfrak{A}$ is an ideal of R, then $S = 1 + \mathfrak{A} = \{1 + x \mid x \in \mathfrak{A}\}$ is another example of a multiplicatively closed subset of R.

Suppose S is a multiplicatively closed subset of R. Suppose $\mathfrak{B}$ is an ideal of R contained in S^c. The ideal $\mathfrak{B}$ is said to be maximal with respect to S if every ideal of R which properly contains $\mathfrak{B}$ has a nontrivial intersection with S. Thus, $\mathfrak{B}$ is maximal with respect to S if $\mathfrak{B} \subseteq S^c$, and $\mathfrak{B} < \mathfrak{A}$ ($\mathfrak{A}$ an ideal of R) implies $\mathfrak{A} \cap S \neq \emptyset$. Another way to say the same thing is $\mathfrak{B}$ is maximal with respect to S if $\mathfrak{B}$ is maximal (with respect to $\subseteq$) among the ideals of R contained in S^c. If $\mathfrak{B}$ is maximal with respect to S, then $\mathfrak{B}$ is a prime ideal of R. In fact, we have the following more general result.

Theorem 6.3 Let S be a multiplicatively closed subset of R. Suppose $\mathfrak{A}$ is an ideal of R contained in S^c. Then there exists an ideal $\mathfrak{P} \subseteq S^c$ such that $\mathfrak{A} \subseteq \mathfrak{P}$ and $\mathfrak{P}$ is maximal with respect to S. Any such $\mathfrak{P}$ is a prime ideal of R.

Proof. Let $\mathcal{F} = \{\mathfrak{B} \mid \mathfrak{A} \subseteq \mathfrak{B} \subseteq S^c$, and $\mathfrak{B}$ an ideal of $R\}$. The set $\mathcal{F}$ is nonempty since $\mathfrak{A} \in \mathcal{F}$. Partially order $\mathcal{F}$ by inclusion $\subseteq$ (see Appendix A). If $\Gamma = \{\mathfrak{A}_\alpha \mid \alpha \in \Delta\}$ is a chain in $(\mathcal{F}, \subseteq)$, then clearly $\cup_{\alpha \in \Delta} \mathfrak{A}_\alpha$ is an upper bound of Γ in $\mathcal{F}$. Thus, we can apply Zorn's lemma and find a maximal element $\mathfrak{P}$ of $\mathcal{F}$. The ideal $\mathfrak{P}$ contains $\mathfrak{A}$ and is maximal with respect to S.

Suppose $\mathfrak{P}$ is any ideal of R which contains $\mathfrak{A}$ and is maximal with respect to S. We claim $\mathfrak{P}$ is a prime ideal of R. To see this, suppose $\mathfrak{P}$ is not prime. Then there exist elements $x,y \in \mathfrak{P}^c$ such that $xy \in \mathfrak{P}$. Since $x \notin \mathfrak{P}$, $Rx + \mathfrak{P} > \mathfrak{P}$. Since $\mathfrak{P}$ is maximal with respect to S, $(Rx + \mathfrak{P}) \cap S \neq \emptyset$. Let $s \in (Rx + \mathfrak{P}) \cap S$. By the same reasoning, there exists an element $s' \in (Ry + \mathfrak{P}) \cap S$. Then $ss' \in S$ since S is multiplicatively closed. Also,

$$ss' \in (Rx + \mathfrak{P})(Ry + \mathfrak{P}) \subseteq Rxy + \mathfrak{P} \subseteq \mathfrak{P}.$$

Therefore, $ss' \in S \cap \mathfrak{P}$. This is impossible since $\mathfrak{P} \subseteq S^c$. We conclude $\mathfrak{P}$ is a prime ideal of R. This completes the proof of Theorem 6.3. ∎

One important application of Theorem 6.3 is the following corollary concerning maximal ideals in R.

Corollary 6.4 Let $\mathfrak{A}$ be a proper ideal of R. Then $\mathfrak{A}$ is contained in a maximal ideal of R. Furthermore, any maximal ideal of R is prime.

Proof Let $S = \{1\}$. Clearly, S is a multiplicatively closed subset of R, and $\mathfrak{A} \subseteq S^c$ since $\mathfrak{A} \neq R$. Theorem 6.3 implies $\mathfrak{A} \subseteq \mathfrak{P}$ for some ideal $\mathfrak{P}$ which is maximal with respect to S. But $\mathfrak{P}$ is maximal with respect to $S = \{1\}$ if and only if $\mathfrak{P}$ is a maximal ideal of R. Hence, $\mathfrak{A}$ is contained in a maximal ideal of R.

Suppose $\mathfrak{P}$ is any maximal ideal of R. Then $(0) \subseteq \mathfrak{P}$, and $\mathfrak{P}$ is maximal with respect to $\{1\}$. Thus, Theorem 6.3 implies $\mathfrak{P}$ is a prime ideal of R. ∎

We next introduce the radical of an ideal.

Definition 6.5 Let $\mathfrak{A}$ be an ideal of R. The radical of $\mathfrak{A}$, denoted by $\sqrt{\mathfrak{A}}$, is the set $\sqrt{\mathfrak{A}} = \{x \in R \mid x^n \in \mathfrak{A} \text{ for some } n \geq 1\}$.

It is easy to see that $\sqrt{\mathfrak{A}}$ is an ideal of R with $\mathfrak{A} \subseteq \sqrt{\mathfrak{A}}$. If $x \mapsto \bar{x}$ denotes the natural map of R onto $R/\mathfrak{A}$, then $\sqrt{\mathfrak{A}} = \{x \in R \mid \bar{x} \text{ is nilpotent in } R/\mathfrak{A}\}$. In particular, $\sqrt{(0)}$ is the nil radical of R.

Theorem 6.6 Let $\mathfrak{A}$ be a proper ideal of R. Then $\sqrt{\mathfrak{A}}$ is the intersection of all prime ideals of R which contain $\mathfrak{A}$.

Proof. We will let $V(\mathfrak{A})$ denote the set of all prime ideals of R which contain $\mathfrak{A}$. By Corollary 6.4, $V(\mathfrak{A}) \neq \emptyset$. Let $\mathfrak{P} \in V(\mathfrak{A})$, and suppose $x \in \sqrt{\mathfrak{A}}$. Then $x^n \in \mathfrak{A} \subseteq \mathfrak{P}$ for some $n \geq 1$. Since $\mathfrak{P}$ is a prime ideal, $x \in \mathfrak{P}$. Therefore, $\sqrt{\mathfrak{A}} \subseteq \cap \{\mathfrak{P} \mid \mathfrak{P} \in V(\mathfrak{A})\}$.

For the converse inclusion, suppose $x \notin \sqrt{\mathfrak{A}}$. Then $S = \{x^i \mid i \geq 0\}$ is a multiplicatively closed subset of R which is disjoint from $\mathfrak{A}$. By Theorem 6.3, there exists a prime ideal $\mathfrak{P}_1 \in V(\mathfrak{A})$ such that $x \notin \mathfrak{P}_1$. Thus, $x \notin \cap \{\mathfrak{P} \mid \mathfrak{P} \in V(\mathfrak{A})\}$. We conclude $\cap \{\mathfrak{P} \mid \mathfrak{P} \in V(\mathfrak{A})\} \subseteq \sqrt{\mathfrak{A}}$. ∎

One special case of Theorem 6.6 is worth mentioning here. If $\mathfrak{A} = (0)$, then $\sqrt{(0)}$ is called the nil radical (or prime radical) of R. As mentioned above, $\sqrt{(0)}$ is just the set of all nilpotent elements of R. Theorem 6.6 implies $\sqrt{(0)}$ is the intersection of all prime ideals of R. By Theorem 1.6, $J(R)$ is the intersection of all maximal ideals of R. Hence, Corollary 6.4 imples $\sqrt{(0)} \subseteq J(R)$. In general these two ideals are different. Consider the following simple example.

Example 6.7 Let p be a prime in $\mathbb{Z}$ with $p \geq 2$. Set $R = \{x/y \in \mathbb{Q} \mid p \not| y\}$. Thus, R consists of those rational numbers x/y (with g.c.d. $(x,y) = 1$) for which y is not divisible by p. The reader can easily check that R is a subring of $\mathbb{Q}$ (see Exercise 2 at the end of this chapter). It is also easy to see that R contains precisely one maximal ideal $\mathfrak{M} = \{x/y \in R \mid p|x\}$. Thus, $J(R) = \mathfrak{M}$. On the other hand, R is an integral domain, and consequently, $\sqrt{(0)} = (0)$. Thus, $\sqrt{(0)} \neq J(R)$. ∎

Let $\mathfrak{A}$ be a proper ideal of R. We can partially order $V(\mathfrak{A})$, the set of all primes in R containing $\mathfrak{A}$, by inclusion $\subseteq$.

Definition 6.8 Let $\mathfrak{A}$ be a proper ideal of R. A prime ideal of R which contains $\mathfrak{A}$ and is minimal with respect to inclusion in $V(\mathfrak{A})$ is called a minimal prime of $\mathfrak{A}$.

Thus, an ideal $\mathfrak{P}$ is a minimal prime of $\mathfrak{A}$ ($\mathfrak{A} \neq R$), if $\mathfrak{P}$ is a prime ideal, $\mathfrak{A} \subseteq \mathfrak{P}$, and there is no prime ideal $\mathfrak{P}'$ of R with $\mathfrak{A} \subseteq \mathfrak{P}' < \mathfrak{P}$. We will show that every ideal different from R has at least one minimal prime. For now, let us observe that if $\mathfrak{A}$ is a prime ideal of R, then $\mathfrak{A}$ is the only minimal prime of $\mathfrak{A}$. Let us consider some other examples of minimal primes.

Example 6.9

(a) Suppose R is a unique factorization domain (e.g., any PID as in 1.7, or a polynomial ring $F[X_1, \ldots, X_r]$ in r variables over a field F). Let $f \in R^*$ and set $\mathfrak{A} = Rf$, the principal ideal generated by f. We assume $f \notin U(R)$. Let $f = p_1^{\alpha(1)} p_2^{\alpha(2)} \cdots p_n^{\alpha(n)}$ be a factorization of f into irreducible factors. Thus, $p_1, \ldots, p_n$ are distinct, nonassociate primes in R, and $\alpha(1), \ldots, \alpha(n) \geq 1$. Since R is a unique factorization domain, each principal ideal Rp_i is a prime ideal of R. It is easy to check that $\{Rp_1, \ldots, Rp_n\}$ is the complete set of minimal primes of $\mathfrak{A}$.

(b) Let $R = F[x,y]$ as in Example 5.24. Then $\mathfrak{M} = (x,y)$ is the only minimal prime of $\mathfrak{A} = Rx$. Notice that $\mathfrak{M}$ is also a maximal ideal of R. Thus, a minimal prime of $\mathfrak{A}$ could also be a maximal ideal of R.

(c) Let $R = F[X,Y]$. Here F is a field, and X, Y are indeterminates over F. Set $\mathfrak{A} = (XY^2, X^2Y)$. The minimal primes of $\mathfrak{A}$ are $\{\mathfrak{P}_1 = RX, \mathfrak{P}_2 = RY\}$. Notice that $\mathfrak{M} = (X,Y)$ is a prime ideal of R which contains $\mathfrak{A}$, but $\mathfrak{M}$ is not a minimal prime of $\mathfrak{A}$. ∎

There is a nice connection between minimal primes of $\mathfrak{A}$ and maximal, multiplicatively closed subsets of R which are disjoint from $\mathfrak{A}$. Suppose S is a multiplicatively closed subset of R which is disjoint from $\mathfrak{A}$ ($\neq R$). Thus, $\mathfrak{A} \subseteq$

S^c. The set S is called a maximal, multiplicatively closed subset of R disjoint from $\mathfrak{A}$ if S has the following property: If T is a multiplicatively closed subset of R with $S < T$, then $T \cap \mathfrak{A} \neq \emptyset$. We can always imbed a multiplicatively closed subset of R which is disjoint from $\mathfrak{A}$ into a maximal, multiplicatively closed subset of R disjoint from $\mathfrak{A}$.

6.10 Let $\mathfrak{A}$ be a proper ideal of R. Suppose S is a multiplicatively closed subset of R disjoint from $\mathfrak{A}$. Then S is contained in a maximal, multiplicatively closed subset of R disjoint from $\mathfrak{A}$.

The statement in 6.10 is a simple consequence of Zorn's lemma. Let

$$\mathfrak{F} = \{T \subseteq R \mid T \text{ is multiplicatively closed, } S \subseteq T, \text{ and } T \cap \mathfrak{A} \doteq \emptyset\}$$

$\mathfrak{F} \neq \emptyset$ since $S \in \mathfrak{F}$. Partially order $\mathfrak{F}$ by inclusion. If $\{T_\alpha \mid \alpha \in \Delta\} = \Gamma$ is a chain in $(\mathfrak{F},\subseteq)$, then $\cup_{\alpha\in\Delta} T_\alpha$ is an upper bound of Γ in $\mathfrak{F}$. Thus, by Zorn's lemma, $\mathfrak{F}$ has a maximal element T. Obviously, T is a maximal, multiplicatively closed subset of R disjoint from $\mathfrak{A}$. Since $S \subseteq T$, the assertion in 6.10 is proved.

The connection between minimal primes of $\mathfrak{A}$ and maximal, multiplicatively closed subsets of R disjoint from $\mathfrak{A}$ is given in our next theorem.

Theorem 6.11 Let $\mathfrak{A}$ be a proper ideal of R. An ideal $\mathfrak{P}$ is a minimal prime of $\mathfrak{A}$ if and only if $\mathfrak{P}^c$ is a maximal, multiplicatively closed subset of R disjoint from $\mathfrak{A}$.

Proof. Suppose $\mathfrak{P}^c$ is a maximal, multiplicatively closed subset of R disjoint from $\mathfrak{A}$. Since $\mathfrak{P}^c$ is a multiplicatively closed subset, $\mathfrak{P}$ is a prime ideal of R. Since $\mathfrak{P}^c \cap \mathfrak{A} = \emptyset$, $\mathfrak{A} \subseteq \mathfrak{P}$. Suppose $\mathfrak{Q}$ is a prime ideal of R such that $\mathfrak{A} \subseteq \mathfrak{Q} \subseteq \mathfrak{P}$. Then $\mathfrak{P}^c \subseteq \mathfrak{Q}^c$, and $\mathfrak{Q}^c$ is a multiplicatively closed subset of R which is disjoint from $\mathfrak{A}$. By the maximality of $\mathfrak{P}^c$, we conclude $\mathfrak{P}^c = \mathfrak{Q}^c$. Therefore, $\mathfrak{Q} = \mathfrak{P}$, and $\mathfrak{P}$ is a minimal prime of $\mathfrak{A}$.

Conversely, suppose $\mathfrak{P}$ is a minimal prime of $\mathfrak{A}$. Then $\mathfrak{P}^c$ is a multiplicatively closed subset of R which is disjoint from $\mathfrak{A}$. By 6.10, $\mathfrak{P}^c \subseteq T$ where T is a maximal, multiplicatively closed subset of R disjoint from $\mathfrak{A}$. By Theorem 6.3 there exists a prime ideal $\mathfrak{Q}$ such that $\mathfrak{A} \subseteq \mathfrak{Q}$, and $\mathfrak{Q} \cap T = \emptyset$. In particular, $\mathfrak{A} \subseteq \mathfrak{Q} \subseteq \mathfrak{P}$. Since $\mathfrak{P}$ is a minimal prime of $\mathfrak{A}$, $\mathfrak{Q} = \mathfrak{P}$. But then $\mathfrak{P}^c = T$, and $\mathfrak{P}^c$ is a maximal, multiplicatively closed subset of R disjoint from $\mathfrak{A}$. ∎

We can repeat a portion of the proof of Theorem 6.11 for the following important corollary.

Corollary 6.12 Let $\mathfrak{A}$ be a proper ideal of R. Any prime ideal of R which contains $\mathfrak{A}$ contains a minimal prime of $\mathfrak{A}$.

Proof. Let $\mathfrak{Q}$ be a prime ideal of R such that $\mathfrak{A} \subseteq \mathfrak{Q}$. Then $\mathfrak{Q}^c$ is a multiplicatively closed subset of R which is disjoint from $\mathfrak{A}$. By 6.10, $\mathfrak{Q}^c$ is contained in a maximal, multiplicatively closed subset T of R which is disjoint from $\mathfrak{A}$. Theorem 6.3 implies there exists a prime ideal $\mathfrak{P}$ such that $\mathfrak{A} \subseteq \mathfrak{P}$, and $\mathfrak{P} \cap T = \varnothing$. But then $T \subseteq \mathfrak{P}^c$, and the maximality of T implies $\mathfrak{P}^c = T$. Theorem 6.11 then implies $\mathfrak{P}$ is a minimal prime of $\mathfrak{A}$. Since $\mathfrak{Q}^c \subseteq T = \mathfrak{P}^c$, $\mathfrak{P} \subseteq \mathfrak{Q}$. ∎

Among other things, Corollary 6.12 and Corollary 6.4 imply that every ideal $\mathfrak{A}$ ($\neq R$) has at least one minimal prime $\mathfrak{P}$. Theorem 6.6 and Corollary 6.12 imply that $\sqrt{\mathfrak{A}}$ is precisely the intersection of the minimal primes of $\mathfrak{A}$. Thus, we have

Corollary 6.13 Let $\mathfrak{A}$ be a proper ideal of R. Then $\sqrt{\mathfrak{A}}$ is the intersection of the minimal primes of $\mathfrak{A}$. ∎

Again, let $\mathfrak{A}$ be an ideal of R which is distinct from R. Besides the minimal primes of $\mathfrak{A}$, we will also be interested in the maximal primes belonging to $\mathfrak{A}$. Since $\mathfrak{A} \neq R$, $R/\mathfrak{A}$ is a nonzero R-module. Consider $Z(R/\mathfrak{A})$, the set of zero divisors of the R-module $R/\mathfrak{A}$. From 1.13, we have $x \in Z(R/\mathfrak{A})$ if and only if there exists a $y \in \mathfrak{A}^c$ such that $xy \in \mathfrak{A}$. Notice that $\mathfrak{A} \subseteq Z(R/\mathfrak{A})$.

Definition 6.14 Let $\mathfrak{A}$ and $\mathfrak{B}$ be ideals of R with $\mathfrak{A} \neq R$. The ideal $\mathfrak{B}$ is said to belong to $\mathfrak{A}$ if $\mathfrak{B} \subseteq Z(R/\mathfrak{A})$.

Sometimes authors use the expressions "$\mathfrak{B}$ is related to $\mathfrak{A}$" or "$\mathfrak{B}$ is associated to $\mathfrak{A}$" in place of $\mathfrak{B}$ belongs to $\mathfrak{A}$.

Definition 6.15 Let $\mathfrak{A}$ be a proper ideal of R. A prime ideal $\mathfrak{P}$ of R is called a maximal prime belonging to $\mathfrak{A}$ if $\mathfrak{P}$ belongs to $\mathfrak{A}$ and $\mathfrak{P}$ is maximal among the ideals belonging to $\mathfrak{A}$.

Thus, $\mathfrak{P}$ is a maximal prime belonging to $\mathfrak{A}$ if $\mathfrak{P} \subseteq Z(R/\mathfrak{A})$ and whenever $\mathfrak{P} \subseteq \mathfrak{B} \subseteq Z(R/\mathfrak{A})$ ($\mathfrak{B}$ an ideal of R), then $\mathfrak{P} = \mathfrak{B}$. We first show maximal prime ideals belonging to $\mathfrak{A}$ exist and always contain $\mathfrak{A}$.

Theorem 6.16 Let $\mathfrak{A}$ be a proper ideal of R. There exists a maximal prime ideal belonging to $\mathfrak{A}$. Furthermore, any maximal prime ideal belonging to $\mathfrak{A}$ contains $\mathfrak{A}$.

Proof. Consider the complement $Z(R/\mathfrak{A})^c$ of $Z(R/\mathfrak{A})$. An element x lies in $Z(R/\mathfrak{A})^c$ if and only if x satisfies the following property: Whenever $xy \in \mathfrak{A}$, then $y \in \mathfrak{A}$. This characterization of $Z(R/\mathfrak{A})^c$ implies $Z(R/\mathfrak{A})^c$ is a multiplicatively closed subset of R. For suppose x_1 and x_2 are elements of $Z(R/\mathfrak{A})^c$. If $(x_1x_2)y$

$\in \mathfrak{A}$ for some $y \in R$, then $x_2 y \in \mathfrak{A}$ since $x_1 \in Z(R/\mathfrak{A})^c$ Since $x_2 \in Z(R/\mathfrak{A})^c$, $y \in \mathfrak{A}$. Therefore, $x_1 x_2 \in Z(R/\mathfrak{A})^c$. Clearly $1 \in Z(R/\mathfrak{A})^c$. Consequently, $Z(R/\mathfrak{A})^c$ is a multiplicatively closed subset of R which is necessarily disjoint from $\mathfrak{A}$ since $\mathfrak{A} \subseteq Z(R/\mathfrak{A})$.

Theorem 6.3 implies there exists a prime ideal $\mathfrak{P}$ with $\mathfrak{A} \subseteq \mathfrak{P}$ and such that $\mathfrak{P}$ is maximal with respect to $Z(R/\mathfrak{A})^c$. Then $\mathfrak{P}$ is a maximal prime ideal belonging to $\mathfrak{A}$.

Finally, suppose $\mathfrak{P}$ is any maximal prime ideal belonging to $\mathfrak{A}$. Then $\mathfrak{P} \subseteq \mathfrak{P} + \mathfrak{A} \subseteq Z(R/\mathfrak{A})$. By the maximality of $\mathfrak{P}$ in $Z(R/\mathfrak{A})$, we conclude $\mathfrak{P} = \mathfrak{P} + \mathfrak{A}$. In particular, $\mathfrak{A} \subseteq \mathfrak{P}$. ∎

If $\mathfrak{P}$ is a maximal prime ideal belonging to $\mathfrak{A}$, then $\mathfrak{P}$ is a prime ideal of R, but $\mathfrak{P}$ need not be a maximal ideal of R. Consider the following example.

Example 6.17 Let R be a unique factorization domain as in Example 6.9a. Suppose $p_1, \ldots, p_n$ are distinct (nonassociate) primes in R. Let $f = p_1^{\alpha(1)} p_2^{\alpha(2)} \cdots p_n^{\alpha(n)}$. Here $\alpha(i) \geq 1$ for each $i = 1, \ldots, n$. We have seen in Example 6.9a that $\{Rp_1, \ldots, Rp_n\}$ is the complete set of minimal primes of Rf.

On the other hand, $x \in Z(R/Rf)$ if and only if $p_i \mid x$ for some $i = 1, \ldots, n$. Therefore, $Z(R/Rf) = Rp_1 \cup Rp_2 \cup \cdots \cup Rp_n$. It easily follows from this (see Exercises 11 and 12 at the end of this chapter) that $\{Rp_1, \ldots, Rp_n\}$ is precisely the set of maximal primes belonging to Rf. If $R = F[X_1, \ldots, X_r]$ with $r \geq 2$, then $Rp_1, \ldots, Rp_n$ are not maximal ideals of R. ∎

In Example 6.17, the minimal primes of $\mathfrak{A}$ are the same as the maximal primes belonging to $\mathfrak{A}$. In general, these two sets of primes are different.

Example 6.18 Let $\mathfrak{A} = (XY^2, X^2Y)$ in Example 6.9c. We have seen that the minimal primes of $\mathfrak{A}$ are $\{RX, RY\}$. It is easy to see that $Z(R/\mathfrak{A}) = \mathfrak{M} = (X,Y)$. Since $\mathfrak{M}$ is a maximal ideal of R, $\{\mathfrak{M}\}$ is the complete set of maximal prime ideals belonging to $\mathfrak{A}$. ∎

Although the set of minimal primes of $\mathfrak{A}$ and the set of maximal primes belonging to $\mathfrak{A}$ are in general different, we do have every minimal prime of $\mathfrak{A}$ is contained in some maximal prime ideal belonging to $\mathfrak{A}$. In order to prove this assertion, we need the following observation. Suppose $\mathfrak{B}$ is an ideal which belongs to $\mathfrak{A}$. Thus, $\mathfrak{B} \subseteq Z(R/\mathfrak{A})$. Then $\mathfrak{B}$ and the multiplicatively closed subset $Z(R/\mathfrak{A})^c$ are disjoint. Theorem 6.3 implies $\mathfrak{B} \subseteq \mathfrak{P}$ where $\mathfrak{P}$ is a prime ideal maximal with respect to $Z(R/\mathfrak{A})^c$. Thus, $\mathfrak{P}$ is a maximal prime ideal belonging to $\mathfrak{A}$. In particular, any ideal belonging to $\mathfrak{A}$ is contained in a maximal prime ideal belonging to $\mathfrak{A}$.

Theorem 6.19 Let $\mathfrak{A}$ be a proper ideal of R. Then every minimal prime of $\mathfrak{A}$ is contained in a maximal prime ideal belonging to $\mathfrak{A}$.

Proof. Let $\mathfrak{P}$ be a minimal prime of $\mathfrak{A}$. Since every ideal belonging to $\mathfrak{A}$ is contained in a maximal prime ideal belonging to $\mathfrak{A}$, it suffices to show $\mathfrak{P} \subseteq Z(R/\mathfrak{A})$. Since $\mathfrak{P} \subseteq Z(R/\mathfrak{A})$ if and only if $Z(R/\mathfrak{A})^c \subseteq \mathfrak{P}^c$, we will show $Z(R/\mathfrak{A})^c \subseteq \mathfrak{P}^c$.

Let $x \in Z(R/\mathfrak{A})^c$. Let S be the following subset of R:

6.20 $S = \{x^i, z, zx^i \mid z \in \mathfrak{P}^c, \text{ and } i \geq 0\}$.

Thus, S consists of three types of elements: powers of x, elements from $\mathfrak{P}^c$, and all products of powers of x with elements from $\mathfrak{P}^c$. Since $\mathfrak{P}^c$ and $\{x^i \mid i \geq 0\}$ are multiplicatively closed subsets of R, it is easy to see S is also a multiplicatively closed subset of R. Clearly, $\mathfrak{P}^c \subseteq S$.

We claim $S \cap \mathfrak{A} = \varnothing$. To see this, we look at each type of element in S. Since $\mathfrak{A} \subseteq \mathfrak{P}$, no element from $\mathfrak{P}^c$ lies in $\mathfrak{A}$. To see that no power of x lies in $\mathfrak{A}$, we need the following observation:

6.21 $\sqrt{\mathfrak{A}} \subseteq Z(R/\mathfrak{A})$.

Suppose $b \in \sqrt{\mathfrak{A}}$. Then $b^n \in \mathfrak{A}$ for some positive integer n. We have observed earlier that $\mathfrak{A} \subseteq Z(R/\mathfrak{A})$. Thus, if $n = 1$, $b \in Z(R/\mathfrak{A})$. Hence, we can assume $b \notin \mathfrak{A}$. Then $n \geq 2$. We can assume n is the smallest positive integer such that $b^n \in \mathfrak{A}$. Thus, $b^{n-1} \notin \mathfrak{A}$. But then $b^n = b(b^{n-1}) \in \mathfrak{A}$ implies $b \in Z(R/\mathfrak{A})$. This proves 6.21.

Now suppose some power of x lies in $\mathfrak{A}$. Say $x^i \in \mathfrak{A}$. Here $i \geq 1$, since $1 \notin \mathfrak{A}$. Then $x \in \sqrt{\mathfrak{A}} \subseteq Z(R/\mathfrak{A})$ by 6.21. This is impossible since x (and all its powers) lie in the multiplicatively closed subset $Z(R/\mathfrak{A})^c$. Thus, no power of x lies in $\mathfrak{A}$.

Finally, suppose $zx^i \in \mathfrak{A}$ for some $i \geq 1$ and some $z \in \mathfrak{P}^c$. Among all such elements from S, we can choose one, say zx^r, with r as small as possible. Since $zx^r \in \mathfrak{A}$, r must be at least 2. (If $zx \in \mathfrak{A}$, then $z \notin \mathfrak{A}$ implies $x \in Z(R/\mathfrak{A})$, which is impossible.) The element $zx^{r-1} \notin \mathfrak{A}$ by the minimality of r. But then $x(zx^{r-1}) = zx^r \in \mathfrak{A}$, and $zx^{r-1} \notin \mathfrak{A}$ implies $x \in Z(R/\mathfrak{A})$, which is impossible. We conclude that no element of S of the form zx^i, $z \in \mathfrak{P}^c$ and $i \geq 0$, can lie in $\mathfrak{A}$. Thus, $S \cap \mathfrak{A} = \varnothing$ as claimed.

Since $\mathfrak{P}$ is a minimal prime of $\mathfrak{A}$, $\mathfrak{P}^c$ is a maximal, multiplicatively closed subset of R disjoint from $\mathfrak{A}$ by Theorem 6.11. Therefore, $\mathfrak{P}^c = S$. In particular, $x \in \mathfrak{P}^c$. Since x is an arbitrary element in $Z(R/\mathfrak{A})^c$, we conclude $Z(R/\mathfrak{A})^c \subseteq \mathfrak{P}^c$. This completes the proof of Theorem 6.19. ∎

EXERCISES

1. Let R be an integral domain. Suppose S is a multiplicatively closed subset of R. We assume $0 \notin S$. Let $Q(R)$ denote the quotient field of R. Set $S^{-1}R = \{x/s \in Q(R) \mid s \in S, x \in R\}$.
   (a) Show $S^{-1}R$ is a subring of $Q(R)$.
   (b) If $\mathfrak{P}$ is a prime ideal of R such that $\mathfrak{P} \cap S = \emptyset$, show $S^{-1}\mathfrak{P} = \{x/s \in S^{-1}R \mid x \in \mathfrak{P}\}$ is a prime ideal of $S^{-1}R$.
   (c) Show that every prime ideal of $S^{-1}R$ is of the form $S^{-1}\mathfrak{P}$ with $\mathfrak{P}$ a prime ideal of R disjoint from S.
2. Show that the ring constructed in Example 6.7 is a special case of Exercise 1 above. Use the information from Exercise 1 to prove all the statements about R in Example 6.7.
3. Let F be a field, and set $R = F[[X]]$, the ring of formal power series in X over F. Determine all prime ideals of R.
4. Let $R = \mathbb{Z}[i]$, the ring of Gaussian integers. Thus, $i^2 = -1$. Determine all prime ideals of R.
5. Let $R = \mathbb{Z}[X,Y]$. Here X and Y are indeterminates. Find a chain of prime ideals in R of length three; i.e., find primes $\mathfrak{P}_1, \mathfrak{P}_2, \mathfrak{P}_3$ of R such that $\mathfrak{P}_1 < \mathfrak{P}_2 < \mathfrak{P}_3$. Are there any longer chains of prime ideals in R?
6. Let R be an arbitrary commutative ring. Show the following sets are multiplicatively closed subsets of R:
   (a) $S =$ any one of the four examples given after Definition 6.2 in the text.
   (b) $S = \cap \{T_\alpha \mid \alpha \in \Delta\}$ where each T_α is a multiplicatively closed subset of R.
   (c) $S = T_1T_2 = \{xy \mid x \in T_1, y \in T_2\}$ where T_1 and T_2 are multiplicatively closed subsets of R.
   (d) $S = \cup \{T_\alpha \mid \alpha \in \Delta\}$ where $\{T_\alpha \mid \alpha \in \Delta\}$ is a collection of multiplicatively closed subsets of R with the following property: For all $\alpha, \beta, \in \Delta$, there exists $\gamma \in \Delta$ such that $T_\alpha \cup T_\beta \subseteq T_\gamma$.
7. Determine the ideals of $\mathbb{Z}$ which are maximal with respect to $S = \{2^i \mid i \geq 0\}$. Do the same problem for $S = 1 + 2\mathbb{Z}$.
8. Verify that $\{Rp_1, \ldots, Rp_n\}$ in Example 6.9a is the complete set of minimal primes of $\mathfrak{A} = Rf$.
9. Let $R = \mathbb{Z}[X,Y]$ as in Exercise 5. Let $\mathfrak{A} = (X^2Y, XY^2)$. Compute the minimal primes of $\mathfrak{A}$.
10. In the proof of Theorem 6.16, we claimed $\mathfrak{P} + \mathfrak{A} \subseteq Z(R/\mathfrak{A})$. Give a proof of this statement.
11. Suppose $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ are distinct prime ideals of R. Let $\mathfrak{A}$ be an ideal of R. Show $\mathfrak{A} \subseteq \cup_{i=1}^n \mathfrak{P}_i$ if and only if $\mathfrak{A}$ is contained in some $\mathfrak{P}_i$.
12. Use Exercise 11 to show $\{Rp_1, \ldots, Rp_n\}$ are the maximal primes belonging to $\mathfrak{A} = Rf$ in Example 6.9a.
13. Find the maximal primes belonging to $\mathfrak{A}$ in Exercise 9.

14. Give an example of a commutative ring R and two ideals $\mathfrak{A}$ and $\mathfrak{B}$ of R such that $\mathfrak{B}$ belongs to $\mathfrak{A}$ but $\mathfrak{A}$ does not belong to $\mathfrak{B}$.
15. In Exercise 14, is such an example possible if $R = \mathbb{Z}$?
16. Find a ring R and ideals $\mathfrak{A}$ and $\mathfrak{P}$ of R such that $\mathfrak{P}$ is prime, $\mathfrak{A} \subseteq \mathfrak{P}$, but $\mathfrak{P}$ is neither a minimal prime of $\mathfrak{A}$ nor a maximal prime belonging to $\mathfrak{A}$.
17. Suppose R is a commutative ring without an identity element 1. Construct an example which shows Corollary 6.4 need not be true for such a ring.

7

The Cayley-Hamilton Theorem

Let F be a field, and suppose $A \in M_{n\times n}(F)$. Let $C_A(X)$ denote the characteristic polynomial of A. Thus, $C_A(X) = \det(XI_n - A)$. The classical Cayley-Hamilton theorem says $C_A(A) = O$. In this chapter, we will show this theorem is valid for any commutative ring R. In proving this result, we will exploit the natural isomorphism $M_{n\times n}(R[X]) \cong (M_{n\times n}(R))[X]$. In particular, we need to say a few words about polynomial rings $T[X]$ where T is not necessarily commutative.

Let T be a ring. We do not assume T is commutative. Let X be an indeterminate over T and consider the polynomial ring

$$T[X] = \{a_0X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n \mid a_i \in T, \text{ and } n \geq 0\}$$

Polynomials in $T[X]$ are added or multiplied together in the usual ways (see [5, Chapter 2]). Thus, $T[X]$ is an associative ring with identity $1 (=$ the identity of $T)$. The elements in T are identified with the constants of $T[X]$. Remember when dealing with $T[X]$, that X is in the center of $T[X]$. Thus, $Xf(X) = f(X)X$ for all $f(X) \in T[X]$. The ring $T[X]$, in general, is not commutative since $(aX)(bX) = (ab)X^2$ for any $a, b \in T$ and $(bX)(aX) = (ba)X^2$. These products are equal if and only if $ab = ba$ in T. Obviously, $T[X]$ is a commutative ring if and only if T is a commutative ring.

Suppose $f(X) \in T[X]^*$. Then there exist a unique integer $n \geq 0$ and unique elements $a_0, \ldots, a_n \in T$ such that $f(X) = a_0X^n + a_1X^{n-1} + \cdots + a_n$ and

$a_0 \neq 0$. The element a_0 is called the leading coefficient of $f(X)$ and the integer n is called the degree of $f(X)$. In this text, we will let $\partial(f)$ denote the degree of $f(X)$. Thus, if $f(X) = a_0X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n \in T[X]$ and $a_0 \neq 0$, then $\partial(f) = n$. Notice the degree defines a function $\partial : T[X]^* \mapsto \mathbb{N}_0$ where $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. The zero polynomial in this chapter is not given a degree. (See Exercise 2 at the end of the chapter.) We have the usual rules concerning the degree function.

7.1 (a) $\partial(f) = 0$ if and only if $f \in T^*$.
(b) If f, g, and $f + g$ are nonzero polynomials in $T[X]$, then $\partial(f + g) \leq \max\{\partial(f), \partial(g)\}$.
(c) If f, g, and fg are nonzero polynomials in $T[X]$, then $\partial(fg) \leq \partial(f) + \partial(g)$.

The most important fact about $T[X]$ used in this text is the following division theorem.

Theorem 7.2 Let $f(X) = a_0X^n + \cdots + a_n$ and $g(X) = b_0X^m + \cdots + b_m$ be polynomials in $T[X]$. Assume $b_0 \in U(T)$. Then there exist unique polynomials $u(X)$, $v(X)$, $r(X)$, and $s(X)$ in $T[X]$ such that
(a) $f(X) = u(X)g(X) + r(X)$, with $r(X) = 0$, *or* $\partial(r) < \partial(g)$ and
(b) $f(X) = g(X)v(X) + s(X)$ with $s(X) = 0$, *or* $\partial(s) < \partial(g)$

Proof. The proofs of the assertions in (a) and (b) are very similar. We will prove (a) and leave (b) to the reader. We first construct polynomials $u(X)$ and $r(X)$ such that (a) is satisfied. We will then argue $u(X)$ and $r(X)$ are necessarily unique.

Since b_0 is a unit in T, $b_0 \neq 0$. In particular, $g(X) \neq 0$, and $\partial(g) = m \geq 0$. If $f(X) = 0$, then $u(X) = r(X) = 0$ satisfy (a). Hence, we can assume $a_0 \neq 0$. Thus, $\partial(f) = n \geq 0$. If $n < m$, then $u(X) = 0$ and $r(X) = f(X)$ satisfy (a). Hence, we can assume $n \geq m$ and proceed by induction on n.

The polynomial $f_1(X) = f(X) - a_0b_0^{-1}X^{n-m}g(X)$ is either zero or a nonzero polynomial of degree less than n. Hence, our induction hypotheses imply there exist polynomials $u_1(X)$ and $r(X)$ in $T[X]$ such that $f_1(X) = u_1(X)g(X) + r(X)$, and $r(X)$ is either zero or $\partial(r) < \partial(g)$. We then have

$$f(X) = (u_1(X) + a_0b_0^{-1}X^{n-m})g(X) + r(X)$$

Set $u(X) = u_1(X) + a_0b_0^{-1}X^{n-m}$. Then $f(X) = u(X)g(X) + r(X)$.

In order to finish the proof of (a), we must argue the polynomials $u(X)$ and $r(X)$ constructed in the last paragraph are unique. Suppose there is another pair

of polynomials $u_1(X)$ and $r_1(X)$ such that $f(X) = u_1(X)g(X) + r_1(X)$ with $r_1(X) = 0$ *or* $\partial(r_1) < \partial(g)$. Then

$$u(X)g(X) + r(X) = f(X) = u_1(X)g(X) + r_1(X)$$

Thus, $(u - u_1)g = r_1 - r$ in $T[X]$. Suppose $u - u_1 \neq 0$. Since b_0 is a unit in T, $(u - u_1)g \neq 0$ and

$$\partial((u - u_1)g) = \partial(u - u_1) + \partial(g) = \partial(u - u_1) + m \geq m$$

On the other hand, either $r_1 - r$ is zero or $\partial(r_1 - r) < \partial(g) = m$. Thus, $(u - u_1)g = r_1 - r$ is impossible. We conclude $u_1(X) = u(X)$. Then $r_1(X) = r(X)$, and the proof of the theorem is complete. ∎

If $f(X) = u(X)g(X) + r(X)$ with $r(X) = 0$ or $\partial(r) < \partial(g)$, we will say g divides f on the right with remainder r. On the other hand, if $f(X) = g(X)v(X) + s(X)$ with $s = 0$ or $\partial(s) < \partial(g)$, then we say g divides f on the left with remainder s. Theorem 7.2 implies the following statement about division: If the leading coefficient b_0 of $g(X) = b_0X^m + \cdots + b_m \in T[X]$ is a unit in T, then $g(X)$ divides any $f(X) \in T[X]$ on the right or left with suitable remainders. In particular, if $g(X)$ is a monic polynomial, i.e., the leading coefficient of $g(X)$ is 1, then $g(X)$ divides any $f(X) \in T[X]$ on the right or left with suitable remainders. Notice also that if the leading coefficient of $g(X)$ is a unit in T, then $g(X)$ is a regular element in $T[X]$, and $\partial(fg) = \partial(f) + \partial(g)$ for any nonzero $f(X) \in T[X]$.

If T is a commutative ring, then the uniqueness argument in the proof of Theorem 7.2 implies $u(X) = v(X)$, and $r(X) = s(X)$ in (a) and (b). If T is not commutative, then $u(X)$ and $v(X)$ (or $r(X)$ and $s(X)$) need not be equal.

Definition 7.3 Let $f(X), g(X) \in T[X]$, and assume $g(X) \neq 0$.

(a) If $f(X) = u(X)g(X) + r(X)$ with $r = 0$, or $\partial(r) < \partial(g)$, then $r(X)$ is called a right remainder of division of $f(X)$ by $g(X)$.

(b) If $f(X) = g(X)v(X) + s(X)$ with $s = 0$, or $\partial(s) < \partial(g)$, then $s(X)$ is called a left remainder of division of $f(X)$ by $g(X)$.

If the leading coefficient of $g(X)$ is a unit in T, then Theorem 7.2 implies a right (left) remainder of division of $f(X)$ by $g(X)$ is unique. However, a right remainder need not be equal to a left remainder of division of $f(X)$ by $g(X)$. Consider the following example.

Example 7.4 Let $T = M_{2\times 2}(\mathbb{Q})$. Set

$$f(X) = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} X + \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \in T[X]$$

Let

$$g(X) = X + \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \in T[X]$$

Notice that $g(X)$ is a monic polynomial in $T[X]$ of degree 1. Then

$$\textbf{7.5} \quad \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\left[X + \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}\right] = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} X + \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}$$
$$u(X) \qquad g(X) \quad = \qquad f(X)$$

Thus,

$$r(X) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

is the right remainder of division of $f(X)$ by $g(X)$. On the other hand, we have

$$\textbf{7.6} \quad \left[X + \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}\right]\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} + \begin{bmatrix} -2 & 4 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} X + \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}$$
$$g(X) \qquad v(X) \quad + \quad s(X) \quad = \qquad f(X)$$

Therefore,

$$s(X) = \begin{bmatrix} -2 & 4 \\ 0 & 2 \end{bmatrix}$$

is the left remainder of division of $f(X)$ by $g(X)$. ∎

Of course, if T is a commutative ring and the leading coefficient of $g(X)$ is a unit in T, then the right (left) remainder of division of $f(X)$ by $g(X)$ is unique and the right remainder is the same as the left remainder.

Suppose R is a commutative ring and

$$f(X) = a_0X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n$$

is a polynomial in $R[X]$. If $z \in R$, then $f(z)$ has an unambiguous meaning, namely

$$f(z) = a_0z^n + a_1z^{n-1} + \cdots + a_{n-1}z + a_n.$$

The map sending $f(X) \mapsto f(z)$ is an R-algebra homomorphism from $R[X]$ to R. If T is a noncommutative ring, then $f(z)$ can have two possible meanings

Definition 7.7 Let

$$f(X) = a_0X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n \in T[X]$$

Let $z \in T$. Then $f_R(z)$ and $f_L(z)$ will denote the following two elements of T:

(a) $f_R(z) = a_0z^n + a_1z^{n-1} + \cdots + a_{n-1}z + a_n$.

(b) $f_L(z) = z^na_0 + z^{n-1}a_1 + \cdots + za_{n-1} + a_n$.

$f_R(z)$ is called the right evaluation of $f(X)$ at z, and $f_L(z)$ is called the left evaluation of $f(X)$ at z. If T is not commutative, then obviously $f_R(z)$ need not be equal to $f_L(z)$. Notice, however, that if $a_0, \ldots, a_{n-1} \in C(T)$, or if $z \in C(T)$, then $f_R(z) = f_L(z)$. In particular, if T is a commutative ring, then $f_R(z) = f(z) = f_L(z)$.

We can now state the noncommutative version of the remainder theorem.

Theorem 7.8 Let $f(X) \in T[X]$, and let $z \in T$. Then there exist polynomials $u(X)$ and $v(X)$ in $T[X]$ such that

(a) $f(X) = u(X)(X - z) + f_R(z)$.

(b) $f(X) = (X - z)v(X) + f_L(z)$.

Thus, $f_R(z)$ is the right remainder of division of $f(X)$ by $X - z$ and $f_L(z)$ is the left remainder of division of $f(X)$ by $X - z$.

Proof. We will prove (a) and leave the proof of (b) to the reader. By Theorem 7.2, there exist unique polynomials $u(X)$ and $r(X)$ in $T[X]$ such that

$$\textbf{7.9} \quad f(X) = u(X)(X - z) + r(X).$$

Furthermore, either $r(X)$ is zero or $\partial(r) < \partial(X - z) = 1$. Thus, $r(X) = r$, some constant in T. If $f(X)$ is zero, or if $\partial(f) = 0$, then $u(X) = 0$ and $f(X) = r = f_R(z)$. Hence, we can assume $\partial(f) = n \geq 1$. Then equation 7.9 implies $\partial(u) = n - 1$. Suppose

$$u(X) = c_0X^{n-1} + c_1X^{n-2} + \cdots + c_{n-2}X + c_{n-1}$$

Substituting this expression in equation 7.9, we have

$$\textbf{7.10} \quad f(X) = c_0X^n + (c_1 - c_0z)X^{n-1} + \cdots + (c_{n-1} - c_{n-2}z)X + (r - c_{n-1}z)$$

Therefore, substituting z on the right in equation 7.10 for X, we have

$$f_R(z) = c_0z^n + (c_1 - c_0z)z^{n-1} + \cdots + (c_{n-1} - c_{n-2}z)z + (r - c_{n-1}z) = r$$

∎

If $f(X) = u(X)v(X)$ in $T[X]$, then $u(X)$ is called a left divisor or left factor of $f(X)$. Similarly, $v(X)$ is called a right divisor or right factor of $f(X)$. Using this terminology, we have the following corollary to Theorem 7.8.

Corollary 7.11 $X - z$ is a right divisor of $f(X)$ in $T[X]$ if and only if $f_R(z) = 0$. Similarly, $X - z$ is a left divisor of $f(X)$ in $T[X]$ if and only if $f_L(z) = 0$. ∎

Now suppose R is a commutative ring. Then $R[X]$ is a commutative ring, and we can consider the noncommutative ring $M_{n\times n}(R[X])$. Let $A \in M_{n\times n}(R[X])$. Then for each $i, j = 1, \ldots, n$, $[A]_{ij}$ is a polynomial in $R[X]$. If $A \neq O$, let p be the maximum of the degrees of the nonzero entries of A. Thus,

$$p = \max\{\partial([A]_{ij}) \mid [A]_{ij} \neq 0,\ 1 \leq i,j \leq n\}.$$

If $A = O$, set $p = 0$. Since p is a fixed, nonnegative integer, each entry $[A]_{ij}$ can be written uniquely as an R-linear combination of $X^p, X^{p-1}, \ldots, X, 1$. Hence, for each $i,j = 1, \ldots, n$, there exist unique elements $a_0^{(i,j)}, \cdots, a_p^{(i,j)} \in R$ such that

$$\textbf{7.12}\quad [A]_{ij} = a_0^{(i,j)}X^p + a_1^{(i,j)}X^{p-1} + \cdots + a_{p-1}^{(i,j)}X + a_p^{(i,j)}$$

Substituting these expressions for the $[A]_{ij}$ in A, we have the following polynomial description of A.

$$\textbf{7.13}\quad A = \begin{bmatrix} a_0^{(1,1)}X^p + \cdots + a_p^{(1,1)}, \ldots, a_0^{(1,n)}X^p + \cdots + a_p^{(1,n)} \\ \vdots \qquad\qquad\qquad \vdots \\ a_0^{(n,1)}X^p + \cdots + a_p^{(n,1)}, \ldots, a_0^{(n,n)}X^p + \cdots + a_p^{(n,n)} \end{bmatrix}$$

$$= \begin{bmatrix} a_0^{(1,1)}, \ldots, a_0^{(1,n)} \\ \vdots \qquad \vdots \\ a_0^{(n,1)}, \ldots, a_0^{(n,n)} \end{bmatrix} X^p + \begin{bmatrix} a_1^{(1,1)}, \ldots, a_1^{(1,n)} \\ \vdots \qquad \vdots \\ a_1^{(n,1)}, \ldots, a_1^{(n,n)} \end{bmatrix} X^{p-1} + \cdots$$

$$+ \begin{bmatrix} a_p^{(1,1)}, \ldots, a_p^{(1,n)} \\ \vdots \qquad \vdots \\ a_p^{(n,1)}, \ldots, a_p^{(n,n)} \end{bmatrix}$$

Thus, every matrix $A \in M_{n\times n}(R[X])$ can be written in the following form:

7.14 $A = A_0X^p + A_1X^{p-1} + \cdots + A_{p-1}X + A_p.$

with $A_0, A_1, \ldots, A_p \in M_{n\times n}(R)$. Since $XB = BX$ for every $B \in M_{n\times n}(R[X])$, we could also write equation 7.14 as

7.15 $A = X^pA_0 + X^{p-1}A_1 + \cdots + XA_{p-1} + A_p.$

We have noted that the coefficients appearing in equation 7.12 are unique. Hence, it follows that the matrices $A_0, \ldots, A_p$ appearing in equation 7.14 or 7.15 are also unique. In fact, it is clear that $\sum_{j=0}^{k} A_jX^{k-j} = \sum_{j=0}^{k} B_jX^{k-j}$, with $A_0, \ldots, A_k$ and $B_0, \ldots, B_k \in M_{n\times n}(R)$ and $k \geq 0$, implies $A_0 = B_0$, $A_1 = B_1, \ldots, A_k = B_k$.

We have now established the following representation principle.

7.16 Every $n\times n$ matrix A with entries from $R[X]$ can be written uniquely in the form

$$A = A_0X^p + A_1X^{p-1} + \cdots + A_{p-1}X + A_p$$

with $A_0, \ldots, A_p \in M_{n\times n}(R)$ and

$$p = \max\{\partial([A]_{ij}) \mid [A]_{ij} \neq 0,\ 1 \leq i,j \leq n\}$$

If $A = O$, p is taken to be 0.
Consider the following example.

Example 7.17 Suppose

$$A = \begin{bmatrix} 2X^2 + 3X + 1 & 6X - 5 \\ 2 & X^2 - 1 \end{bmatrix} \in M_{2\times 2}(\mathbb{Z}[X])$$

Then

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} X^2 + \begin{bmatrix} 3 & 6 \\ 0 & 0 \end{bmatrix} X + \begin{bmatrix} 1 & -5 \\ 2 & -1 \end{bmatrix}$$ ∎

Now let $T = M_{n\times n}(R)$. We can consider the (noncommutative) polynomial ring $T[X] = (M_{n\times n}(R))[X]$. The representation given in 7.16 suggests a natural mapping $\psi : M_{n\times n}(R[X]) \mapsto (M_{n\times n}(R))[X]$ given by $\psi(A) = \sum_{j=0}^{p} A_jX^{p-j}$. The reader can easily check that ψ is an isomorphism of rings. Hence, we have the following lemma:

Lemma 7.18 The rings $M_{n\times n}(R[X])$ and $(M_{n\times n}(R))[X]$ are isomorphic via the map $\psi(A) = \sum_{j=0}^{p} A_jX^{p-j}$ (notation as in 7.16). ∎

We will make use of the isomorphism $\psi : M_{n\times n}(R[X]) \cong (M_{n\times n}(R))[X]$ when proving the Cayley-Hamilton theorem.

Definition 7.19 Let $A \in M_{n\times n}(R)$. The characteristic polynomial of A, written $C_A(X)$, is defined as follows: $C_A(X) = \det(XI_n - A)$.

Expanding the determinant of $XI_n - A$ using Laplace's equation, we see $C_A(X)$ is a polynomial in $R[X]$ of the following form:

$$\textbf{7.20}\quad C_A(X) = X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n.$$

In equation 7.20, $a_1 = -\sum_{j=1}^{n} [A]_{jj} = -Tr(A)$ and $a_n = (-1)^n \det(A)$. The other coefficients in the characteristic polynomial have various interpretations, but they are not relevant for this discussion. The important thing to notice here is that the characteristic polynomial of any matrix $A \in M_{n\times n}(R)$ is always a monic polynomial of degree n in $R[X]$. In particular, $C_A(X)$ is a regular element of $R[X]$ for any matrix A.

Every monic polynomial of degree n in $R[X]$ is the characteristic polynomial of some $n\times n$ in $M_{n\times n}(R)$. To see this, suppose

$$f(X) = X^n + c_1X^{n-1} + \cdots + c_{n-1}X + c_n$$

is a monic polynomial in $R[X]$. Using Laplace's expansion, the reader can easily verify

$$\textbf{7.21}\quad \det\begin{bmatrix} X & 0 & 0 & 0 & \cdots & 0 & c_n \\ -1 & X & 0 & 0 & \cdots & 0 & c_{n-1} \\ 0 & -1 & X & 0 & \cdots & 0 & c_{n-2} \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & -1 & X + c_1 \end{bmatrix} = f(X)$$

The $n\times n$ matrix whose determinant appears in equation 7.21 can be written as $XI_n - \mathrm{Com}(f)$, where $\mathrm{Com}(f)$ is the following $n\times n$ matrix from $M_{n\times n}(R)$:

$$\textbf{7.22}\quad \mathrm{Com}(f) = \begin{bmatrix} 0 & 0 & & 0 & -c_n \\ 1 & 0 & \cdots & 0 & -c_{n-1} \\ 0 & 1 & \cdots & 0 & -c_{n-2} \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & & 1 & -c_1 \end{bmatrix}$$

The $n \times n$ matrix $\mathrm{Com}(f)$ appearing in equation 7.22 is called the companion matrix of $f(X)$. Equation 7.21 implies the characteristic polynomial of the companion matrix of $f(X)$ is $f(X)$. In particular, every monic polynomial of degree n in $R[X]$ is the characteristic polynomial of some matrix from $M_{n\times n}(R)$.

The Cayley-Hamilton theorem says A is a zero of its characteristic polynomial.

Theorem 7.23 (Cayley-Hamilton) Let $A \in M_{n\times n}(R)$. Then $C_A(A) = \mathrm{O}$.

Proof. Since $C_A(X) = \det(XI_n - A)$, equation 2.20 implies

$$\mathbf{7.24} \quad \mathrm{adj}(XI_n - A)(XI_n - A) = C_A(X)I_n.$$

Equation 7.24 is a statement about matrices in $M_{n\times n}(R[X])$. Using Lemma 7.18, we can reinterpret equation 7.24 in the polynomial ring $(M_{n\times n}(R))[X]$. Suppose $C_A(X) = X^n + a_1X^{n-1} + \cdots + a_n$. Let

$$f(X) = X^n + (a_1I_n)X^{n-1} + \cdots + (a_nI_n) \in (M_{n\times n}(R))[X]$$

Let ψ denote the ring isomorphism from $M_{n\times n}(R[X])$ onto $(M_{n\times n}(R))[X]$ given in Lemma 7.18. Then $\psi(C_A(X)I_n) = f(X)$, and $\psi(XI_n - A) = XI_n - A$ $(= X - A)$, the corresponding linear polynomial in $(M_{n\times n}(R))[X])$. Applying ψ to equation 7.24, we see $X - A$ is a right divisor of $f(X)$ in $(M_{n\times n}(R))[X]$. In particular, $f_R(A) = \mathrm{O}$ by Corollary 7.11. Therefore

$$\mathrm{O} = f_R(A) = A^n + (a_1I_n)A^{n-1} + \cdots + (a_nI_n) = C_A(A)I_n.$$

We conclude $C_A(A) = \mathrm{O}$. ∎

There are many applications of Theorem 7.23. For instance, if $A \in \mathrm{Gl}(n,R)$, then the Cayley-Hamilton theorem implies the inverse of A must be a polynomial in A.

Corollary 7.25 Let $A \in \mathrm{Gl}(n,R)$. Then $A^{-1} = g(A)$ for some $g(X) \in R[X]$.

Proof. If $n = 1$, then $A = (a)$ with $a \in U(R)$. We can take $g(X) = a^{-1} \in R \subseteq R[X]$. Then $g(A) = a^{-1}I_1 = (a^{-1}) = A^{-1}$. Hence, we can assume $n > 1$.

Let $C_A(X) = X^n + a_1X^{n-1} + \cdots + a_n$. We have noted that $a_n = (-1)^n \det(A)$. Since A is invertible, a_n is a unit in R by Corollary 2.21. By Theorem 7.23, $C_A(A) = \mathrm{O}$. Therefore,

$$\mathbf{7.26} \quad A[(-a_n^{-1})(A^{n-1} + a_1A^{n-2} + \cdots + a_{n-1}I_n)] = I_n$$

Set $g(X) = -a_n^{-1}X^{n-1} - a_n^{-1}a_1X^{n-2} + \cdots + (-a_n^{-1}a_{n-1}) \in R[X]$. Then equation 7.26 implies $g(A) = A^{-1}$ ∎

Let $A \in M_{n\times n}(R)$. The matrix A determines an R-algebra homomorphism $\vartheta_A : R[X] \mapsto M_{n\times n}(R)$ given by the following formula.

$$\textbf{7.27}\quad \vartheta_A(r_0X^m + r_1X^{m-1} + \cdots + r_{m-1}X + r_m) = r_0A^m + r_1A^{m-1} + \cdots + r_{m-1}A + r_mI_n.$$

It is easy to see that ϑ_A is an R-algebra homomorphism from $R[X]$ to $M_{n\times n}(R)$. Notice that ϑ_A sends the identity 1 in $R[X]$ to the identity I_n of $M_{n\times n}(R)$. The image of ϑ_A is the subring of $M_{n\times n}(R)$ generated by R and A. We will denote this subring by $R[A]$. Clearly, $R[A]$ is set of all polynomials in A with coefficients from R. Thus,

$$\mathrm{Im}(\vartheta_A) = R[A] = \{f(A) \mid f(X) \in R[X]\}$$

Notice that $R[A]$ is a commutative subring of $M_{n\times n}(R)$. The kernel of ϑ_A gets a formal name in this discussion.

Definition 7.28 Let $A \in M_{n\times n}(R)$. The kernel of the R-algebra homomorphism $\vartheta_A : R[X] \mapsto M_{n\times n}(R)$ is called the null ideal (or characteristic ideal) of A.

We will denote the null ideal of A by N_A. Do not confuse the null ideal of A with the null space of A. The null space of A is the R-submodule of R^n given by $\mathrm{NS}(A) = \{\xi \in R^n \mid A\xi = \mathrm{O}\}$. The null ideal of A is the ideal in $R[X]$ given by $N_A = \mathrm{Ker}(\vartheta_A) = \{f(X) \in R[X] \mid f(A) = \mathrm{O}\}$

A sequence

$$0 \mapsto P \overset{g}{\mapsto} M \overset{f}{\mapsto} Q \mapsto 0$$

of R-modules P, M, and Q and R-module homomorphisms g and f is said to be exact if g is injective, f is surjective, and $\mathrm{Im}(g) = \mathrm{Ker}(f)$. If we let $\iota : N_A \mapsto R[X]$ denote the inclusion map of N_A into $R[X]$, then the definitions imply

$$\textbf{7.29}\quad 0 \mapsto N_A \overset{\iota}{\mapsto} R[X] \overset{\vartheta_A}{\mapsto} R[A] \mapsto 0$$

is an exact sequence of R-modules.

The Cayley-Hamilton theorem says $C_A(X) \in N_A$. In particular, the null ideal of a matrix is never zero. In fact, N_A always contains a regular element (namely $C_A(X)$) of $R[X]$. In the classical case, i.e., when $R = F$ a field, the ring $F[X]$ is a PID. Consequently, N_A is a principal ideal in $F[X]$. The unique monic polynomial in $F[X]$ which generates N_A is called the minimal polynomial of A. For an arbitrary commutative ring R, the ring $R[X]$ need not be a PID and N_A need not be principal. Consider the following example.

Example 7.30 Let $R = \mathbb{Z}/4\mathbb{Z} = \{0,1,2,3\}$. Let

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \in M_{2\times 2}(R)$$

Then

$$C_A(X) = (X - 1)^2 = X^2 - 2X + 1 = X^2 + 2X + 1 = (X + 1)^2$$

Thus, $(X + 1)^2 \in N_A$. Since $2(A + I_2) = O$, we also have $2(X + 1) \in N_A$. A simple calculation shows $2(X + 1)$ is the only linear polynomial in N_A.

Suppose $f(X) \in N_A$. By Theorem 7.2, there exist unique polynomials $u(X)$ and $r(X)$ in $R[X]$ such that $f(X) = u(X)(X + 1)^2 + r(X)$. Furthermore, either $r(X)$ is zero or $\partial(r) < 2$. Evaluating this expression at A implies $r(A) = 0$. Thus, $r = 0$ or $r = 2(X + 1)$. We have now shown $N_A = ((X + 1)^2, 2(X + 1))$.

We claim the ideal $N_A = ((X + 1)^2, 2(X + 1))$ is not a principal ideal in $R[X]$. To see this, we assume N_A is principal and derive a contradiction. In order to make computations easier, consider the following change of variables. Let $\sigma : R[X] \cong R[X]$ be the isomorphism given by $\sigma(X) = X + 3$ and $\sigma(y) = y$ for all $y \in R$. $\sigma(N_A) = (X^2,2X)$. Thus, if N_A is a principal ideal in $R[X]$, then $(X^2,2X)$ is a principal ideal in $R[X]$.

Suppose $(X^2,2X) = (p)$ for some $p(X) \in R[X]$. Since $p = aX^2 + b(2X)$ for some $a,b \in R[X]$, $X|p$. Write $p(X) = Xp_1(X)$. Then $X(X,2) = X(p_1)$. Since X is a regular element in $R[X]$, we conclude $(X,2) = (p_1)$. We will show that $(X,2)$ is not a principal ideal in $R[X]$, and thus we have a contradiction.

In $R[X]/2R[X] \cong \mathbb{Z}/2\mathbb{Z}[X]$, the image $\bar{p}_1$ of p_1 generates the principal ideal (X). Thus, $\bar{p}_1 \sim X$ in $\mathbb{Z}/2\mathbb{Z}[X]$. The only unit in the ring $\mathbb{Z}/2\mathbb{Z}[X]$ is 1. Thus, $p_1 = X$. Therefore, $p_1(X) = X + 2g(X)$ for some $g(X) \in R[X]$. Hence, we can write $p_1(X)$ in the following form:

$$p_1(X) = 2a_0 + (1 + 2a_1)X + 2a_2X^2 + \cdots + 2a_nX^n$$

Here $a_0, \ldots, a_n \in R = \mathbb{Z}/4\mathbb{Z}$. Since 2 is nilpotent in R, $1 + 2a_1$ is a unit in R. Replacing $p_1(X)$ with $(1 + 2a_1)^{-1}p_1(X)$ if need be, we can assume $p_1(X) = 2a_0 + X + 2a_2X^2 + \cdots + 2a_nX^n$.

Now $2a_0 = 0$ or 2 for any $a_0 \in R$. Suppose $2a_0 = 0$. Since $(X,2) = (p_1)$,

$$2 = f(X)p_1(X) = f(X)(X + 2a_2X^2 + \cdots + 2a_nX^n)$$

for some $f(X) \in R[X]$. This is clearly impossible. Hence, $2a_0 = 2$. But then

$$2 = (2 + X + 2a_2X^2 + \cdots + 2a_nX^n)(b_0 + b_1X + \cdots + b_mX^m)$$

for some $b_0, \ldots, b_m \in R$. In particular, $2b_0 = 2$ and $2b_1 + b_0 = 0$. The reader can easily check these two equations have no common solution b_0 and b_1 in R. Hence, in either case, $2a_0 = 0$ or $2a_0 = 2$, we get a contradiction. We conclude

that $(X,2)$ is not principal in $R[X]$, and, consequently, N_A is not principal in $R[X]$. ∎

Example 7.30 shows that when R is not a field, N_A can contain some unexpected polynomials. If we replace $\mathbb{Z}/4\mathbb{Z}$ with $\mathbb{Q}$ in Example 7.30, then N_A is principal, generated by $C_A(X) = X^2 - 2X + 1$. In particular, N_A contains no linear polynomial. If R is not a field, it may be difficult to compute N_A for a given matrix A. In our next theorem, we give necessary and sufficient conditions for a polynomial to lie in the null ideal of a matrix.

Theorem 7.31 Let $A \in M_{n\times n}(R)$. Let $g(X) \in R[X]$. Then $g(X) \in N_A$ if and only if $g(X)\ \mathrm{adj}(XI_n - A) = KC_A(X)$ for some $K \in M_{n\times n}(R[X])$.

Proof. Suppose there exists an $n \times n$ matrix $K \in M_{n\times n}(R[X])$ such that

$$\textbf{7.32}\quad g(X)\ \mathrm{adj}(XI_n - A) = KC_A(X).$$

Since $R[X]I_n$ lies in the center of the ring $M_{n\times n}(R[X])$, $C_A(X)$ can appear on either side of K in equation 7.32. Multiplying equation 7.32 on the right with $XI_n - A$ and using 2.20, we have

$$\textbf{7.33}\quad C_A(X)g(X)I_n = C_A(X)K(XI_n - A).$$

Via the canonical isomorphism $\psi : M_{n\times n}(R[X]) \cong (M_{n\times n}(R))[X]$, we can view equation 7.33 as a statement about polynomials in $(M_{n\times n}(R))[X]$. Since $C_A(X)$ is a monic polynomial, $\psi(C_A(X)I_n)$ is a regular element in $(M_{n\times n}(R))[X]$. Thus, $\psi(C_A(X)I_n)$ can be canceled after applying ψ to equation 7.33. We then have

$$\textbf{7.34}\quad \psi(g(X)I_n) = \psi(K)(X - A) \quad \text{in } (M_{n\times n}(R))[X]$$

Suppose

$$g(X) = b_0X^m + b_1X^{m-1} + \cdots + b_{m-1}X + b_m$$

Here $b_0, \ldots, b_m$ are elements in R. Then

$$\psi(g(X)I_n) = (b_0I_n)X^m + (b_1I_n)X^{m-1} + \cdots + (b_{m-1}I_n)X + b_mI_n \quad \text{in } (M_{n\times n}(R))[X]$$

Set $\psi(g(X)I_n) = P(X)$. The coefficients of $P(X)$, namely $b_0I_n, \ldots, b_mI_n$, lie in the center of the ring $M_{n\times n}(R)$. Consequently, $P_L(A) = P_R(A)$. Equation 7.34 implies $X - A$ is a right divisor of $P(X)$. Hence, $P_R(A) = P_L(A) = \mathrm{O}$ by Corollary 7.11. Therefore

$$\textbf{7.35}\quad \mathrm{O} = P_R(A) = (b_0I_n)A^m + (b_1I_n)A^{m-1} + \cdots + (b_{m-1}I_n)A + (b_mI_n) = b_0A^m + b_1A^{m-1} + \cdots b_{m-1}A + b_mI_n = g(A)$$

Equation 7.35 implies $g(X) \in N_A$.

Conversely, suppose $g(X) \in N_A$. Then $g(A) = O$. Using the same notation as in equation 7.35, we have $P_R(A) = O$. Thus, by Corollary 7.11, $X - A$ is a right divisor of $P(X)$ in $(M_{n\times n}(R))[X]$. Therefore, $\psi(g(X)I_n) = P(X) = k(X)(X - A)$ for some $k(X)$ in $(M_{n\times n}(R))[X]$. Applying ψ^{-1} to this relationship, we get $g(X)I_n = K(XI_n - A)$ in $M_{n\times n}(R[X])$ for some matrix K. If we now multiply on the right with $\text{adj}(XI_n - A)$, we get equation 7.32. This completes the proof of Theorem 7.31. ∎

Theorem 7.31 implies a polynomial $g(X)$ lies in the null ideal of a matrix A if and only if $C_A(X)$ divides $g(X)$ times each entry of the adjoint of $XI_n - A$. In symbols, $g(X) \in N_A$ if and only if $C_A(X) \mid g(X)\Delta$ for all $\Delta \in I_{n-1}(XI_n - A)$. Let us return to Example 7.30 for an illustration of this result.

Example 7.36 Let $R = \mathbb{Z}/4\mathbb{Z}$, and set

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \in M_{2\times 2}(R)$$

Then

$$XI_2 - A = \begin{bmatrix} X - 1 & -2 \\ 0 & X - 1 \end{bmatrix} = \begin{bmatrix} X + 3 & 2 \\ 0 & X + 3 \end{bmatrix}$$

The nonzero, 1×1 minors of $XI_2 - A$ are $X + 3$ and 2. Theorem 7.31 implies $g(X) \in N_A$ if and only if $C_A(X) \mid (X + 3)g(X)$ and $C_A(X) \mid 2g(X)$. Here $C_A(X) = X^2 + 2X + 1$.

For example, we saw in Example 7.30 that $2(X + 1) \in N_A$. Since

$$(X + 3)2(X + 1) = 2X^2 + 2 = 2C_A(X),$$

$C_A(X) \mid (X + 3)2(X + 1)$. Also, $2(2(X + 1)) = 0$, and therefore $C_A(X) \mid 2(2(X + 1))$. ∎

There are two important corollaries to Theorem 7.31.

Corollary 7.37 Let $A \in M_{n\times n}(R)$. Then $\sqrt{N_A} = \sqrt{(C_A(X))}$.

Proof. $C_A(X) \in N_A$ by the Cayley-Hamilton theorem. Therefore, $\sqrt{(C_A(X))} \subseteq \sqrt{N_A}$. For the other inclusion, let $g(X) \in N_A$. Theorem 7.31 implies $g(X)\,\text{adj}(XI_n - A) = C_A(X)K$ for some $K \in M_{n\times n}(R[X])$. We have seen in the proof of Theorem 7.31 that this last equation implies $\psi(g(X)I_n) = \psi(K)(X - A)$ in $(M_{n\times n}(R))[X]$. Applying ψ^{-1}, we conclude $g(X)I_n = K(XI_n - A)$ in $M_{n\times n}(R[X])$. Taking determinants, we get

$$(g(X))^n = \det(K(XI_n - A)) = \det(K)\det(XI_n - A) = \det(K)C_A(X)$$

Therefore, $g(X) \in \sqrt{(C_A(X))}$. It easily follows from this that $\sqrt{N_A} \subseteq \sqrt{(C_A(X))}$. ∎

Corollary 7.38 Let $A \in M_{n\times n}(R)$. Then the minimal primes of N_A are precisely the same as the minimal primes of $(C_A(X))$.

Proof. Let $\mathfrak{P}$ be a minimal prime of N_A. Then $\sqrt{N_A} \subseteq \mathfrak{P}$. Corollary 7.37 implies $(C_A(X)) \subseteq \mathfrak{P}$. By Corollary 6.12, $\mathfrak{P}$ contains a minimal prime $\mathfrak{Q}$ of $(C_A(X))$. Since $\sqrt{(C_A(X))} \subseteq \mathfrak{Q}$, $N_A \subseteq \mathfrak{Q}$ again by Corollary 7.37. Thus, Corollary 6.12 implies $\mathfrak{Q}$ contains some minimal prime $\mathfrak{P}_1$ of N_A. We now have $N_A \subseteq \mathfrak{P}_1 \subseteq \mathfrak{Q} \subseteq \mathfrak{P}$. Since $\mathfrak{P}$ is a minimal prime of N_A, we conclude $\mathfrak{P} = \mathfrak{P}_1$. In particular, $\mathfrak{P} = \mathfrak{Q}$. Thus, $\mathfrak{P}$ is a minimal prime of $(C_A(X))$. If we reverse the roles of N_A and $(C_A(X))$ in this proof, we get every minimal prime of $(C_A(X))$ is a minimal prime of N_A ∎

Corollary 7.38 is a familiar result when R is a field. If $R = F$, a field, then $F[X]$ is a PID. Thus, N_A is a principal ideal generated by the minimal polynomial $m_A(X)$ of A. We have seen in Example 6.9a, the minimal primes of N_A are $(p_1), \ldots, (p_r)$ where $p_1(X), \ldots, p_r(X)$ are the distinct (nonassociate) irreducible factors of $m_A(X)$. Similarly, the minimal primes of $(C_A(X))$ are $(q_1), \ldots, (q_t)$ where $q_1(X), \ldots, q_t(X)$ are the distinct (nonassociate) irreducible factors of $C_A(X)$. Corollary 7.38 implies $r = t$, and $p_i \sim q_i$ (after possibly reordering $p_1, \ldots, p_r$) for all $i = 1, \ldots, r$. Thus, we recover the following classical result.

Corollary 7.39 Let $A \in M_{n\times n}(F)$ with F a field. Then the irreducible factors of the minimal polynomial of A are exactly the same as the irreducible factors of the characteristic polynomial of A. ∎

The multiplicity of a given irreducible factor in $m_A(X)$ could well be different from the multiplicity of the corresponding factor in $C_A(X)$. For example, if $A = I_n \in M_{n\times n}(F)$, then $m_A(X) = X - 1$ and $C_A(X) = (X - 1)^n$.

In Chapter 9 of this book, we will show the maximal prime ideals belonging to N_A are exactly the same as the maximal prime ideals belonging to $(C_A(X))$. Example 7.30 shows that, in general, $N_A \neq (C_A(X))$. Nevertheless, these two ideals are closely related. They have the same radical, the same minimal primes, and the same maximal primes belonging to each other.

EXERCISES

1. Give examples where the inequalities in 7.1 can be strict.
2. In some textbooks, the zero polynomial is given a degree in the following way. Set $\partial(0) = -\infty$, where $-\infty$ is a symbol satisfying the usual condi-

tions: $-\infty < n$ for all integers n, $(-\infty) + (-\infty) = -\infty$, and $-\infty + n = -\infty$ for all integers n. With this definition for $\partial(0)$, prove the following analog of 7.1:

(a) $\partial(f) \leq 0$ if and only if $f \in T$.

(b) $\partial(f) = 0$ if and only if $f \in T^*$.

(c) $\partial(f + g) \leq \max\{\partial(f),\partial(g)\}$ for any $f,g \in T[X]$.

(d) $\partial(fg) \leq \partial(f) + \partial(g)$ for any $f,g \in T[X]$.

3. Suppose R is an integral domain. Show $U(R[X]) = U(R)$.
4. Give an example which shows Exercise 3 is not true for arbitrary commutative rings.
5. Show $f(X) \in Z(R[X])$ if and only if $cf(X) = 0$ for some nonzero $c \in R$.
6. Suppose R is an integral domain and $f(X) \in R[X]^*$. If $\partial(f) = n$, show $f(X)$ can have at most n roots in R.
7. Exercise 6 is not true in general. Show $X^3 - X$ has six roots in $\mathbb{Z}/6\mathbb{Z}$.
8. Let $\mathbb{H}$ be the quaternions (Exercise 11, Chapter 3). Show $X^2 + 1$ has infinitely many roots in $\mathbb{H}$.
9. Give proofs of Theorems 7.2b and 7.8b.
10. Prove the following variation of Theorem 7.2: Let $f(X),g(X) \in T[X]$. Assume $g(X) \neq 0$, $\partial(g) = m$, and b_0 is the leading coefficient of g. Then there exist an integer $k \geq 0$ and polynomials $q(X)$ and $r(X)$ in $T[X]$ such that $b_0^k f(X) = q(X)g(X) + r(X)$. Furthermore, $r(X) = 0$, or $\partial(r) < m$.
11. Suppose R is a commutative ring. In Definition 7.3, is the remainder unique if the leading coefficient of $g(X)$ is not a unit in R?
12. Let $A \in M_{n\times n}(R)$. Suppose

$$C_A(X) = X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n$$

Show a_i is $(-1)^i$ times the sum of the i-rowed principal ($=$ diagonal) minors of A.

13. Suppose M is an R-module, and let $T \in \mathrm{Hom}_R(M,M)$. Show M is an $R[X]$-module via $f(X)m = f(T)(m)$.
14. Let $f(X)$ be a monic polynomial in $F[X]$. Here F is a field. Show the minimal polynomial of $\mathrm{Com}(f)$ is $f(X)$.
15. Let $A \in M_{n\times n}(R)$. Let X and Y be indeterminates over R. Show $X - Y \mid C_A(X) - C_A(Y)$ in $R[X,Y]$.
16. Give a careful proof of Lemma 7.18.
17. Suppose $f(X)$ is a monic polynomial in $R[X]$. Show $f(X)I_n$ is a regular element in $M_{n\times n}(R[X])$ by using the map ψ.
18. Let $R[[X]]$ denote the ring of formal power series in X over R. Show $M_{n\times n}(R[[X]]) \cong (M_{n\times n}(R))[[X]]$.
19. Compute the null ideals of the following matrices:

(a) $A = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix} \in M_{2\times 2}(\mathbb{Z}/6\mathbb{Z})$

(b) $A = \begin{bmatrix} 3 & 6 \\ 0 & 9 \end{bmatrix} \in M_{2\times 2}(\mathbb{Z}/12\mathbb{Z})$

20. Prove the following theorems of N. McCoy:
Let $A \in M_{n\times n}(R)$. Let $h(X) \in R[X]$. Show $h(A) \in \mathrm{Gl}(n,R)$ if and only if $(h) + N_A = R[X]$.
21. Let $A \in M_{n\times n}(R)$. Let $C'_A(X)$ denote the formal derivative of the characteristic polynomial $C_A(X)$ of A. Show $(C'_A(X)) + N_A = R[X]$ implies $R[A]$ are the only matrices in $M_{n\times n}(R)$ which commute with A.

8
Resultants

In this chapter, we will define resultants and discuss some of their elementary properties. As we will see in the next chapter, resultants can be used to describe a very nice result about zero divisors in $M_{n\times n}(R)$.

Before proceeding with the main topic in this chapter, we will review some well-known definitions about polynomial rings. Let R be a commutative ring, and suppose $X_1, \ldots, X_n$ are indeterminates over R. Consider the polynomial ring $R[X_1, \ldots, X_n]$ (see [5, Chapter 2]). Let $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$. A monic monomial in the variables $X_1, \ldots, X_n$ is any polynomial in $R[X_1, \ldots, X_n]$ of the form $X_1^{\alpha(1)}X_2^{\alpha(2)} \cdots X_n^{\alpha(n)}$. Here $\alpha(1), \ldots, \alpha(n) \in \mathbb{N}_0$. We have mentioned in Example 1.11b that the set of all monic monomials in $R[X_1, \ldots, X_n]$ is a free R-module basis of $R[X_1, \ldots, X_n]$. Thus,

$$\Gamma = \{X_1^{\alpha(1)}X_2^{\alpha(2)} \cdots X_n^{\alpha(n)} \mid (\alpha(1), \ldots, \alpha(n)) \in N_0^n\}$$

is a free R-module basis of $R[X_1, \ldots, X_n]$. Notice $1 = X_1^0X_2^0 \cdots X_n^0$, $X_1 = X_1^1X_2^0 \cdots X_n^0, \ldots, X_n = X_1^0 \cdots X_{n-1}^0X_n^1$ are all elements in Γ. We will refer to Γ as the canonical basis of $R[X_1, \ldots, X_n]$.

Since Γ is a free R-module basis of $R[X_1, \ldots, X_n]$, we have the following property: If $f(X_1, \ldots, X_n)$ is any nonzero polynomial in $R[X_1, \ldots, X_n]$, there exist distinct monomials $M_1, \ldots, M_t \in \Gamma$ and nonzero elements $r_1, \ldots, r_t \in R$ such that $f = r_1M_1 + \cdots + r_tM_t$. Furthermore, such a representation

is unique. If $f = y_1P_1 + \cdots + y_sP_s$ with $P_1, \ldots, P_s$ (distinct) $\in \Gamma$ and $y_1, \ldots, y_s \in R^*$, then $t = s$ and (after a suitable permutation of the P_i) $M_1 = P_1, \ldots, M_t = P_t$ and $r_1 = y_1, \ldots, r_t = y_t$.

There is another representation of the elements in $R[X_1, \ldots, X_n]$ in terms of the free basis Γ, which is often used when studying polynomials. Suppose $f(X_1, \ldots, X_n) \in R[X_1, \ldots, X_n]$. We can think of $f(X_1, \ldots, X_n)$ as a formal power series in $X_1, \ldots, X_n$ with at most finitely many coefficients nonzero. Thus, we can write f in the following form:

$$\textbf{8.1} \quad f(X_1, \ldots, X_n) = \sum c_{\alpha(1)\alpha(2)\cdots\alpha(n)} X_1^{\alpha(1)} X_2^{\alpha(2)} \cdots X_n^{\alpha(n)}$$

In equation 8.1, the sum is taken over all possible n-tuples $(\alpha(1), \ldots, \alpha(n)) \in \mathbb{N}_0^n$. The coefficients $c_{\alpha(1)\alpha(2)\cdots\alpha(n)}$ are all elements of R, and we assume $c_{\alpha(1)\alpha(2)\cdots\alpha(n)} = 0$ except possibly for finitely many indices $(\alpha(1), \ldots, \alpha(n)) \in \mathbb{N}_0^n$. Thus, equation 8.1 is just different notation for the idea that f is a finite linear combination of monomials from Γ. The representation given in equation 8.1 is also unique. If

$$\sum d_{\alpha(1)\alpha(2)\cdots\alpha(n)} X_1^{\alpha(1)} X_2^{\alpha(2)} \cdots X_n^{\alpha(n)} = \sum c_{\alpha(1)\alpha(2)\cdots\alpha(n)} X_1^{\alpha(1)} X_2^{\alpha(2)} \cdots X_n^{\alpha(n)}$$

then $d_{\alpha(1)\alpha(2)\cdots\alpha(n)} = c_{\alpha(1)\alpha(2)\cdots\alpha(n)}$ for all $(\alpha(1), \alpha(2), \ldots, \alpha(n)) \in \mathbb{N}_0^n$. We will use either notation, $f = r_1M_1 + \cdots + r_tM_t$ $(M_1, \ldots, M_t \in \Gamma)$ or that in equation 8.1, whenever it is convenient to do so.

We can now introduce the usual definitions for degree and homogeneous polynomials.

Definition 8.2 Let $f(X_1, \ldots, X_n) \in R[X_1, \ldots, X_n]^*$. Write f as in equation 8.1.

(a) The degree of f, written $\partial(f)$, is the following integer:

$$\partial(f) = \max\{\alpha(1) + \cdots + \alpha(n) \mid c_{\alpha(1)\alpha(2)\cdots\alpha(n)} \neq 0\}.$$

(b) f is a homogeneous polynomial of degree d if

$$\alpha(1) + \cdots + \alpha(n) = d \text{ whenever } c_{\alpha(1)\alpha(2)\cdots\alpha(n)} \neq 0.$$

Notice that any monomial $X_1^{\alpha(1)} X_2^{\alpha(2)} \cdots X_n^{\alpha(n)} \in \Gamma$ has degree $\alpha(1) + \alpha(2) + \cdots + \alpha(n)$. In particular, each monic monomial $X_1^{\alpha(1)} X_2^{\alpha(2)} \cdots X_n^{\alpha(n)}$ is a homogeneous polynomial of degree $\alpha(1) + \cdots + \alpha(n)$. We can extend the definition of degree to the zero polynomial by setting $\partial(0) = -\infty$ with the usual

provisos given in Exercise 2 of Chapter 7. We then have the analog of 7.1 in $R[X_1, \ldots, X_n]$.

8.3 For all $f, g, \in R[X_1, \ldots, X_n]$
(a) $\partial(f) \leq 0$ if and only if $f \in R$.
(b) $\partial(f) = 0$ if and only if $f \in R^*$.
(c) $\partial(f + g) \leq \max\{\partial(f), \partial(g)\}$.
(d) $\partial(fg) \leq \partial(f) + \partial(g)$.

Consider the following examples.

Example 8.4 In $R[X_1, \ldots, X_n]$, assume $n \geq 2$.
(a) Any element in R^* is a homogeneous polynomial of degree 0.
(b) $X_1, \ldots, X_n$ are homogeneous polynomials of degree 1.
(c) $X_1^2X_2^2 + X_2^4 + X_1^3X_2$ is a homogeneous polynomial of degree 4.
(d) $1 + X_1 + X_1X_2 + X_2^3$ is a polynomial of degree 3 but is not homogeneous. ∎

A polynomial $f(X_1, \ldots, X_n) \in R[X_1, \ldots, X_n]$ is homogeneous of degree d if $f = r_1M_1 + r_2M_2 + \cdots + r_tM_t$ with $r_1, \ldots, r_t \in R^*$ and $M_1, \ldots, M_t$ monic monomials from Γ all of degree d. We will soon encounter polynomials for which the representation $f = r_1M_1 + \cdots + r_tM_t$ is not readily apparent. Hence, we present a simple criterion for testing whether a given polynomial is homogeneous or not.

Lemma 8.5 Let $X_1, \ldots, X_n$ and Y be indeterminates over R. Let $f(X_1, \ldots, X_n)$ be a nonzero polynomial in $R[X_1, \ldots, X_n]$. Then f is homogeneous of degree d if and only if $f(YX_1, \ldots, YX_n) = Y^df(X_1, \ldots, X_n)$ in $R[X_1, \ldots, X_n, Y]$.

Proof. Suppose f is homogeneous of degree d. Then $f = r_1M_1 + \cdots + r_tM_t$ for some $r_1, \ldots, r_t \in R^*$ and some $M_1, \ldots, M_t$ all of degree d in Γ. If $M = X_1^{\alpha(1)}X_2^{\alpha(2)} \cdots X_n^{\alpha(n)}$ with $\alpha(1) + \cdots + \alpha(n) = d$, then

$$\begin{aligned} M(YX_1, \ldots, YX_n) &= (YX_1)^{\alpha(1)}(YX_2)^{\alpha(2)} \cdots (YX_n)^{\alpha(n)} \\ &= Y^dX_1^{\alpha(1)} \cdots X_n^{\alpha(n)} = Y^dM(X_1, \ldots, X_n) \end{aligned}$$

It easily follows from this that $f(YX_1, \ldots, YX_n) = Y^df(X_1, \ldots, X_n)$.

Conversely, suppose $f(YX_1, \ldots, YX_n) = Y^df(X_1, \ldots, X_n)$ in $R[X_1, \ldots, X_n, Y]$. Suppose we write f as in equation 8.1. By grouping together the monomials which have the same degree, we can write $f = f_0 + f_1 + \cdots + f_m$ where

each f_i here is either zero or a homogeneous polynomial of degree i. We can assume $f_m \neq 0$ and $m = \partial(f)$. Then

$$Y^d f(X_1, \ldots, X_n) = f(YX_1, \ldots, YX_n) = \sum_{j=0}^{m} f_j(YX_1, \ldots, YX_n)$$
$$= \sum_{j=0}^{m} Y^j f_j(X_1, \ldots, X_n)$$

Since Y is algebraically independent over $R[X_1, \ldots, X_n]$, we conclude $m = d$, and $f_j = 0$ for all $j = 0, \ldots, d-1$. Thus, $f = f_d$, a homogeneous polynomial of degree d. ∎

Suppose S is a commutative R-algebra. Let $f(X_1, \ldots, X_n) \in R[X_1, \ldots, X_n]$. If $s_1, \ldots, s_n \in S$, then we can form $f(s_1, \ldots, s_n) \in S$. If f is given as in equation 8.1, then $f(s_1, \ldots, s_n) = \sum c_{\alpha(1)\alpha(2)\cdots\alpha(n)} s_1^{\alpha(1)} s_2^{\alpha(2)} \cdots s_n^{\alpha(n)}$. The element $f(s_1, \ldots, s_n)$ is called a specialization of $f(X_1, \ldots, X_n)$ in S. Thus, a specialization of $f(X_1, \ldots, X_n)$ in S is just the ring element (in S) obtained by replacing $X_1, \ldots, X_n$ with $s_1, \ldots, s_n$. If we keep $s_1, \ldots, s_n$ fixed, then the map $f(X_1, \ldots, X_n) \mapsto f(s_1, \ldots, s_n)$ is easily seen to be an R-algebra homomorphism from $R[X_1, \ldots, X_n]$ to S. Let us denote this specialization map by $\psi(s_1, \ldots, s_n)$. Thus, $\psi(s_1, \ldots, s_n) : R[X_1, \ldots, X_n] \mapsto S$ is the R-algebra homomorphism given by

$$\psi(s_1, \ldots, s_n)(f(X_1, \ldots, X_n)) = f(s_1, \ldots, s_n).$$

The image of $\psi(s_1, \ldots, s_n)$ is clearly the R-subalgebra of S generated by $\{s_1, \ldots, s_n\}$. We will denote this R-subalgebra by $R[s_1, \ldots, s_n]$. Thus, $\mathrm{Im}(\psi(s_1, \ldots, s_n)) = R[s_1, \ldots, s_n] \subseteq S$.

Any commutative ring R is automatically a (commutative) $\mathbb{Z}$-algebra. Hence, the above remarks about specializations apply in this case. If $r_1, \ldots, r_n \in R$, then $\psi(r_1, \ldots, r_n) : \mathbb{Z}[X_1, \ldots, X_n] \mapsto R$ is a well-defined $\mathbb{Z}$-algebra homomorphism whose image is the subring of R generated by $r_1, \ldots, r_n$. We will need the following result, which is called the Specialization Lemma.

Lemma 8.6 (Specialization Lemma) Let $f(X_1, \ldots, X_n), g(X_1, \ldots, X_n) \in \mathbb{Z}[X_1, \ldots, X_n]$. Suppose $f(j_1, \ldots, j_n) = g(j_1, \ldots, j_n)$ for all positive integers $j_1, \ldots, j_n$. Then for any commutative ring R and any $r_1, \ldots, r_n \in R$, $f(r_1, \ldots, r_n) = g(r_1, \ldots, r_n)$.

Proof. Set

$$F(X_1, \ldots, X_n) = f(X_1, \ldots, X_n) - g(X_1, \ldots, X_n).$$

Then $F(j_1, \ldots, j_n) = 0$ for all positive integers $j_1, \ldots, j_n$. We will show $F(X_1, \ldots, X_n) = 0$. Then for any commutative ring R and any choice of elements $r_1, \ldots, r_n \in R$,

$$\psi(r_1, \ldots, r_n)(F) = \psi(r_1, \ldots, r_n)(0) = 0.$$

In particular, $f(r_1, \ldots, r_n) = g(r_1, \ldots, r_n)$.

We proceed by induction on n. Suppose $n = 1$. Then $F(X_1) \in \mathbb{Z}[X_1]$ has infinitely many roots in $\mathbb{Z}$. By Exercise 6 of Chapter 7, $F(X_1) = 0$.

Suppose $n > 1$. If $\partial(F) \leq 0$, then $F = 0$ since F has infinitely many roots in $\mathbb{Z}$. Suppose $\partial(F) \geq 1$. Then the expansion of F in 8.1 contains at least one nonzero monomial of degree at least one. We can suppose with no loss of generality that X_n appears in this monomial with a positive power. We can then write F as a polynomial in X_n with coefficients in $\mathbb{Z}[X_1, \ldots, X_{n-1}]$.

$$\textbf{8.7} \quad F(X_1, \ldots, X_n) = \sum_{i=0}^{p} G_i(X_1, \ldots, X_{n-1}) X_n^i$$

In equation 8.7, $G_i(X_1, \ldots, X_{n-1}) \in \mathbb{Z}[X_1, \ldots, X_{n-1}]$ for all $i = 0, \ldots, p$ and $G_p(X_1, \ldots, X_{n-1}) \neq 0$. Thus, p is the degree of F when viewed as a polynomial in X_n with coefficients in $\mathbb{Z}[X_1, \ldots, X_{n-1}]$. Since $F(j_1, \ldots, j_n) = 0$ for all positive integers $j_1, \ldots, j_n$, every $G_i(j_1, \ldots, j_{n-1}) = 0$ for all positive integers $j_1, \ldots, j_{n-1}$. For if there is some $i \in \{0, \ldots, p\}$ and some choice of positive integers $j_1, \ldots, j_{n-1}$ for which $G_i(j_1, \ldots, j_{n-1}) \neq 0$, then

$$F(j_1, \ldots, j_{n-1}, X_n) = \sum_{i=0}^{p} G_i(j_1, \ldots, j_{n-1}) X_n^i$$

is a nonzero polynomial in $\mathbb{Z}[X_n]$ with infinitely many roots in $\mathbb{Z}$. We have seen that this is impossible. Thus, for each $i = 0, \ldots, p$, $G_i(j_1, \ldots, j_{n-1}) = 0$ for all positive integers $j_1, \ldots, j_{n-1}$. Our induction hypothesis then implies $G_i(X_1, \ldots, X_{n-1}) = 0$ for all $i = 0, \ldots, p$. In particular, $G_p(X_1, \ldots, X_{n-1}) = 0$. This is impossible. We conclude $\partial(F) \leq 0$. Therefore, $F = 0$. ∎

The Specialization Lemma has many applications in commutative ring theory. An often used refinement of Lemma 8.6 is the following statement: Let $F(X_1, \ldots, X_n) \in \mathbb{Z}[X_1, \ldots, X_n]$. Then $F(r_1, \ldots, r_n) = 0$ for all commutative rings R and all $r_1, \ldots, r_n \in R$ if and only if $F(j_1, \ldots, j_n) = 0$ for all $(j_1, \ldots, j_n) \in \Lambda^n$. Here Λ is any infinite subset of $\mathbb{N}_0$. We leave the proof of this remark to the exercises at the end of this chapter.

We can now introduce Sylvester's matrix and the theory of resultants.

Definition 8.8 Let $u_0, \ldots, u_n$ and $v_0, \ldots, v_m$ be $m + n + 2$ indeterminates over $\mathbb{Z}$. We will always assume $m + n \geq 1$. The following $(m + n) \times (m + n)$ matrix in $M_{(m+n)\times(m+n)}(\mathbb{Z}[u_0, \ldots, u_n, v_0, \ldots, v_m])$ is called Sylvester's matrix

$$\begin{array}{c} \begin{matrix} 1 & 2 & \cdots & n+1 & \cdots & n+m \end{matrix} \\ \left[\begin{array}{ccccccc} u_0 & u_1 & \cdots & u_n & 0 & \cdots & 0 \\ 0 & u_0 & \cdots & u_{n-1} & u_n & \cdots & 0 \\ \vdots & \vdots & & & & & \vdots \\ 0 & 0 & \cdots & u_0 & u_1 & \cdots & u_n \\ v_0 & v_1 & \cdots & v_m & 0 & \cdots & 0 \\ 0 & v_0 & v_1 & \cdots & v_{m-1} & v_m & \cdots & 0 \\ \vdots & \vdots & & & & & \vdots \\ 0 & \cdots & v_0 & v_1 & \cdots & v_m \end{array}\right] \end{array} \begin{matrix} 1 \\ 2 \\ \vdots \\ m \\ m+1 \\ m+2 \\ \vdots \\ m+n \end{matrix}$$

We will let $\mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m)$ denote Sylvester's matrix pictured above. The numbers above $\mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m)$ indicate column numbers and the numbers on the right indicate row numbers. The picture given in Definition 8.8 is meant to symbolize the following construction of $\mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m) = \mathcal{S}$: In the first row of $\mathcal{S}$ is the sequence of variables $u_0, \ldots, u_n$. In the second row of $\mathcal{S}$, is the same sequence $u_0, \ldots, u_n$, only shifted one to the right. This pattern is repeated with $m - 1$ shifts (all blank spaces filled with zeros) to form the first m rows of $\mathcal{S}$:

$$\left[\begin{array}{ccccccccc} u_0 & u_1 & \cdots & u_n & 0 & \cdots & 0 \\ 0 & u_0 & u_1 & \cdots & u_{n-1} & u_n & 0 & \cdots & 0 \\ \vdots & \vdots & & & & & & & \vdots \\ 0 & 0 & \cdots & u_0 & u_1 & \cdots & u_n \end{array}\right] \begin{matrix} 1 \\ 2 \\ \vdots \\ m \end{matrix} \quad (m \times (m+n) \text{ matrix})$$

The $m + 1$ row of $\mathcal{S}$ begins with $v_0, \ldots, v_m$ and then zeros. In each successive row this sequence is shifted one to the right as in the first pattern. This shift is done $n - 1$ times. The last n rows of $\mathcal{S}$ look like

$$\left[\begin{array}{ccccccc} v_0 & v_1 & \cdots & v_m & 0 & \cdots & 0 \\ 0 & v_0 & v_1 & \cdots & v_{m-1} & v_m & \cdots & 0 \\ \vdots & \vdots & & & & & \vdots \\ 0 & \cdots & 0 & v_0 & v_1 & \cdots & v_m \end{array}\right] \begin{matrix} m+1 \\ m+2 \\ \vdots \\ m+n \end{matrix} \quad (n \times (m+n) \text{ matrix})$$

An exact formula for the i,jth entry of $\mathcal{S}$ is given as follows:

$$\textbf{8.9}\quad [\mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m)]_{ij} = \begin{cases} u_{j-i} & \text{for } i = 1, \ldots, m \\ v_{m+j-i} & \text{for } i = m+1, \ldots, m+n \end{cases}$$

In equation 8.9, $u_k = 0$ if $k < 0$ or $k > n$. Similarly, $v_k = 0$ if $k < 0$ or $k > m$.

Notice that the diagonal entries of the (square) matrix $\mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m)$ are $[\mathcal{S}]_{ii} = u_0$ for $i = 1, \ldots, m$ and $[\mathcal{S}]_{ii} = v_m$ for $i = m+1, \ldots, m+n$.

The special cases $n = 0$, $m \geq 1$, and $n \geq 1$, $m = 0$ are also of some interest here. If $n = 0$ and $m \geq 1$, then $\mathcal{S}(u_0, v_0, \ldots, v_m)$ is the $m \times m$ diagonal matrix $\text{Diag}(u_0, \ldots, u_0)$. If $n \geq 1$ and $m = 0$, then $\mathcal{S}(u_0, \ldots, u_n, v_0)$ is the $n \times n$ diagonal matrix $\text{Diag}(v_0, \ldots, v_0)$.

Let us consider some examples.

Example 8.10

(a) $\mathcal{S}(u_0, v_0, v_1) = (u_0)$, and $\mathcal{S}(u_0, u_1, v_0) = (v_0)$.

(b) $\mathcal{S}(u_0,u_1,v_0,v_1) = \begin{bmatrix} u_0 & u_1 \\ v_0 & v_1 \end{bmatrix}$

(c) $\mathcal{S}(u_0,u_1,u_2,v_0,v_1,v_2) = \begin{bmatrix} u_0 & u_1 & u_2 & 0 \\ 0 & u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 & 0 \\ 0 & v_0 & v_1 & v_2 \end{bmatrix}$

(d) $\mathcal{S}(u_0u_1,u_2,v_0,v_1,v_2,v_3) = \begin{bmatrix} u_0 & u_1 & u_2 & 0 & 0 \\ 0 & u_0 & u_1 & u_2 & 0 \\ 0 & 0 & u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 & v_3 & 0 \\ 0 & v_0 & v_1 & v_2 & v_3 \end{bmatrix}$ ∎

Definition 8.11 The determinant $\det(\mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m))$ of Sylvester's matrix is called Sylvester's determinant.

We will let $S(u_0, \ldots, u_n, v_0, \ldots, v_m)$ denote Sylvester's determinant. Thus, $S(u_0, \ldots, u_n, v_0, \ldots, v_m) = \det(\mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m))$. Clearly, $S(u_0, \ldots, u_n, v_0, \ldots, v_m) \in \mathbb{Z}[u_0, \ldots, u_n, v_0, \ldots, v_m]$. In fact, $S(u_0, \ldots, u_n, v_0, \ldots, v_m)$ is always a nonzero polynomial in $u_0, \ldots, u_n, v_0, \ldots, v_m$. To see this, expand $\det(\mathcal{S})$ using Definition 2.16. The diagonal of $\mathcal{S}$ contributes the term $u_0^m v_m^n$ to the sum for $S(u_0, \ldots, u_n, v_0, \ldots, v_m)$. It

is easy to see this is the only monomial in the expansion of $\det(\mathcal{S})$ in which u_0^m and v_m^n both occur. In particular, no other terms in the expansion of $\det(\mathcal{S})$ cancel with $u_0^m v_m^n$. Therefore, $S(u_0, \ldots, u_n, v_0, \ldots, v_m) \neq 0$.

Specializations of Sylvester's determinant are called resultants. Let R be an arbitrary commutative ring.

Definition 8.12 Let

$$f(X) = a_0X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n$$

and

$$g(X) = b_0X^m + b_1X^{m-1} + \cdots + b_{m-1}X + b_m$$

be two polynomials in $R[X]$. We assumed $a_0, b_0 \neq 0$, and $m + n \geq 1$. The resultant of f and g, written $\mathcal{R}(f, g)$, is the following specialization of Sylvester's determinant: $\mathcal{R}(f, g) = S(a_0, \ldots, a_n, b_0, \ldots, b_m)$.

Thus, if $\mathcal{S}(a_0, \ldots, a_n, b_0, \ldots, b_m)$ denotes the $(m + n) \times (m + n)$ matrix in $M_{(m+n)\times(m+n)}(R)$ obtained from $\mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m)$ by replacing $u_0, \ldots, u_n$ with $a_0, \ldots, a_n$ and $v_0, \ldots, v_m$ with $b_0, \ldots, b_m$, respectively, then $\mathcal{R}(f, g) = \det(\mathcal{S}(a_0, \ldots, a_n, b_0, \ldots, b_m))$. Clearly, $\mathcal{R}(f, g) \in R$. Notice that in this text the resultant of f and g is not defined if either f or g is zero or if f and g are both constants in R. Consider the following example.

Example 8.13 Let $R = \mathbb{Z}$. Suppose $f(X) = 2X^3 - 5X^2 + X + 2$ and $g(X) = X^2 - 3X + 2$. Then $n = 3$ and $m = 2$.

$$\mathcal{R}(f,g) = S(2,-5,1,2,1,-3,2) = \det \begin{bmatrix} 2 & -5 & 1 & 2 & 0 \\ 0 & 2 & -5 & 1 & 2 \\ 1 & -3 & 2 & 0 & 0 \\ 0 & 1 & -3 & 2 & 0 \\ 0 & 0 & 1 & -3 & 2 \end{bmatrix} = 0. \quad \blacksquare$$

The most important application of resultants is the following theorem.

Theorem 8.14 Suppose $f(X)$ is a monic polynomial of positive degree in $R[X]$. Let $g(X) \in R[X]^*$. Then $g \in Z(R[X]/(f))$ if and only if $\mathcal{R}(f, g) \in Z(R)$.

The reader will recall from Chapter 6 that g is a zero divisor in $R[X]/(f)$ if and only if the principal ideal (g) belongs to (f). Thus, Theorem 8.14 implies (g) belongs to (f) if and only if $\mathcal{R}(f, g)$ is a zero divisor in R.

Proof of 8.14. Suppose $g(X)$ is a zero divisor of $R[X]/(f)$. Then there exists a polynomial $h(X) \in R[X] - (f)$ such that $h(X)g(X) \in (f(X))$. Let $n = \partial(f)$

and $m = \partial(g)$. Then $n \geq 1$ and $m \geq 0$. By Theorem 7.2, $h(X) = q(X)f(X) + r(X)$ with $q(X)$, $r(X) \in R[X]$ and $r = 0$ or $\partial(r) < n$. Since $h \notin (f)$, $r \neq 0$. Since $h(X)g(X) \in (f)$, there exists a polynomial $k(X) \in R[X]$ such that

8.15 $g(X)r(X) = k(X)f(X)$.

Suppose $k(X)$ is not zero. Since $f(X)$ is monic, equation 8.15 implies

$$n + \partial(k) = \partial(kf) = \partial(rg) \leq \partial(r) + \partial(g) < n + m$$

Thus, $\partial(k) \leq m - 1$. Therefore, $k(X)$ is either zero or a polynomial of degree at most $m - 1$. In particular, we can write the four polynomials f, g, r, and k in the following form:

8.16
$$\begin{aligned} f(X) &= X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n \\ g(X) &= b_0X^m + b_1X^{m-1} + \cdots + b_{m-1}X + b_m \\ r(X) &= c_0X^{n-1} + c_1X^{n-2} + \cdots + c_{n-2}X + c_{n-1} \\ k(X) &= -(d_0X^{m-1} + d_1X^{m-2} + \cdots + d_{m-2}X + d_{m-1}) \end{aligned}$$

In equation 8.16, the coefficients a_i, b_i, c_i, and d_i are elements of R. The element b_0 is not zero since $\partial(g) = m$. The polynomial $r(X)$ is not zero, and $\partial(r) \leq n - 1$. Thus, some c_i is not zero ($c_0 = 0$ if and only if $\partial(r) < n - 1$). The polynomial $k(X)$ may or may not be zero. Thus, we allow the possibility that every d_i is zero in 8.16. Notice the negative sign in the expression for $k(X)$. Substituting these expressions in equation 8.15, we get the following equation.

8.17
$$\begin{aligned} &(b_0X^m + \cdots + b_m)(c_0X^{n-1} + \cdots + c_{n-1}) \\ &\quad + (d_0X^{m-1} + \cdots + d_{m-1})(X^n + \cdots + a_n) = 0 \end{aligned}$$

We can collect coefficients of like powers of X in equation 8.17. Thus, equation 8.17 implies the following equations must all be zero.

8.18
$$\begin{aligned} &b_0c_0 + d_0 = 0 && (m + n - 1) \\ &b_0c_1 + b_1c_0 + a_1d_0 + d_1 = 0 && (m + n - 2) \\ &b_0c_2 + b_1c_1 + b_2c_0 + a_2d_0 + a_1d_1 + d_2 = 0 && (m + n - 3) \\ &\qquad\qquad\vdots \\ &b_{m-1}c_{n-1} + b_mc_{n-2} + a_nd_{m-2} + a_{n-1}d_{m-1} = 0 && (1) \\ &b_mc_{n-1} + a_nd_{m-1} = 0 && (0) \end{aligned}$$

The terms on the right in parentheses in equation 8.18 are the powers of X corresponding to the equation of coefficients, which must be zero. We can view

the equations listed in 8.18 as $m + n$ linear equations in the unknowns $c_0, \ldots, c_{n-1}, d_0, \ldots, d_{m-1}$. In matrix form, these questions can be written as the following homogeneous system of equations.

8.19 $(d_0, \ldots, d_{m-1}, c_0, \ldots, c_{n-1})$
$\mathcal{S}(1, a_1, \ldots, a_n\, b_0, \ldots, b_m) = O$

In equation 8.19, $\mathcal{S}(1, a_1, \ldots, a_n, b_0, \ldots, b_m)$ is Sylvester's matrix with $u_0, \ldots, u_n$ replaced with $1, a_1, \ldots, a_n$ and $v_0, \ldots, v_m$ replaced by $b_0, \ldots, b_m$. Note that $\mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m)$ is well defined since $n + m = \partial(f) + \partial(g) \geq 1$. Taking transposes, we have

8.20 $\mathcal{S}(1, a_1, \ldots, a_n, b_0, \ldots, b_m)^t \xi = O$

In equation 8.20,

$$\xi = (d_0, \ldots, d_{m-1}, c_0, \ldots, c_{n-1})^t \in R^{m+n}$$

Since some c_i here is not zero, ξ is a nontrivial solution to the homogeneous system of equations in 8.20. It follows from Theorem 5.3 and Corollary 4.11d that $\det(\mathcal{S}(1, a_1, \ldots, a_n, b_0, \ldots, b_m)^t) \in Z(R)$. But

$$\det(\mathcal{S}(1, a_1, \ldots, a_n, b_0, \ldots, b_m)^t) = \det(\mathcal{S}(1, a_1, \ldots, a_n, b_0, \ldots, b_m)) = \mathcal{R}(f, g)$$

Therefore, $\mathcal{R}(f, g) \in Z(R)$.

The steps in this argument are easily reversed. If $\mathcal{R}(f, g) \in Z(R)$, then equation 8.20 has a nontrivial solution $\xi \in R^{m+n}$. In particular, the equations in 8.18 have a nontrivial solution $(d_0, \ldots, d_{m-1}, c_0, \ldots, c_{n-1})$. Thus, $g(X)r(X) = k(X)f(X)$ with $r(X)$ and $k(X)$ defined as in equation 8.16. If $c_0 = c_1 = \cdots = c_{n-1} = 0$, then $r(X) = 0$. Then $k(X)f(X) = 0$. Since $f(X)$ is a monic polynomial, we conclude $k(X) = 0$. But then $(d_0, \ldots, d_{m-1}, c_0, \ldots, c_{n-1}) = (0, \ldots, 0)$, which is not the case. Hence, some c_i must be nonzero. Thus, $r(X) \neq 0$. Since $\partial(r) \leq n - 1$, $r(X) \notin (f)$. Therefore, $g(X) \in Z(R[X]/(f))$ ∎

Theorem 8.14 has a particularly nice statement when R is a field.

Corollary 8.21 Suppose F is a field. Let f, g be two polynomials in $F[X]$ of positive degree. Then f and g have a common irreducible factor in $F[X]$ if and only if $\mathcal{R}(f, g) = 0$.

Proof Since F is a field, both f and g are associates of monic polynomials. Hence, with no loss of generality, we can assume both f and g are monic polynomials. Also, $Z(F) = (0)$, since F is a domain.

Suppose $f(X)$ and $g(X)$ have a common irreducible factor $h(X)$ in $F[X]$. Then $\partial(f), \partial(g) \geq \partial(h) \geq 1$. Suppose $f(X) = h(X)f_1(X)$ and $g(X) = h(X)g_1(X)$. Since $\partial(h) \geq 1$, $\partial(f_1) < \partial(f)$. Therefore, $f_1(X) \notin (f)$. We have

$$f_1(X)g(X) = f_1(X)h(X)g_1(X) = g_1(X)f(X) \in (f)$$

Therefore, $g(X) \in Z(F[X]/(f))$. Theorem 8.14 then implies $\mathcal{R}(f, g) = 0$.

Conversely, suppose $\mathcal{R}(f,g) = 0$. Theorem 8.14 implies $g(X) \in Z(F[X]/(f))$. Thus, $g(X)r(X) = k(X)f(X)$ for some polynomials $r(X)$, $k(X) \in F[X]$, and $r(X) \notin (f)$. Now $F[X]$ is a unique factorization domain. If f and g have no common irreducible factor, then $gr = kf$ implies $f \mid r$. Thus, $r(X) \in (f)$. This is impossible. We conclude that f and g must have a common irreducible factor. ∎

There is a slight variant of Corollary 8.21 which we will need in the last part of this chapter.

Corollary 8.22 Suppose R is a unique factorization domain. Let $f(X)$ and $g(X)$ be two polynomials in $R[X]$ of positive degree. Then f and g have a common irreducible factor of positive degree if and only if $\mathcal{R}(f, g) = 0$.

We leave the proof of this corollary as an exercise at the end of this chapter. There is another version of Corollary 8.21 which is used often in algebra.

Corollary 8.23 Let F be a field, and suppose $\overline{F}$ is an algebraic closure of F. Let $f(X)$ and $g(X)$ be two polynomials in $F[X]$ of positive degree. Then f and g have a common root in $\overline{F}$ if and only if $\mathcal{R}(f, g) = 0$.

Proof. The two polynomials $f(X)$ and $g(X)$ have a common root $\alpha \in \overline{F}$ if and only if $X - \alpha$ is a common factor of $f(X)$ and $g(X)$ in $\overline{F}[X]$. Since $\overline{F}$ is algebraically closed, the only irreducible polynomials in $\overline{F}[X]$ are linear polynomials. Hence, f and g have a common root in $\overline{F}$ if and only if $f(X)$ and $g(X)$ have a common irreducible factor in $\overline{F}[X]$. By Corollary 8.21, $f(X)$ and $g(X)$ have a common irreducible factor in $\overline{F}[X]$ if and only if $\mathcal{R}(f, g) = 0$. ∎

Let us return to Example 8.13.

Example 8.24 Let $f(X) = 2X^3 - 5X^2 + X + 2$ and $g(X) = X^2 - 3X + 2$ in $\mathbb{Q}[X]$. We saw in Example 8.13 that $\mathcal{R}(f, g) = 0$. Therefore $f(X)$ and $g(X)$ have a common root in $\overline{\mathbb{Q}}$. It is easy to see that 1 and 2 are both roots of f and g. ∎

In the rest of this chapter, we will explore various properties of Sylvester's determinant. We have already noted that $S(u_0, \ldots, u_n, v_0, \ldots, v_m)$ is a nonzero

polynomial in $\mathbb{Z}[u_0, \ldots, u_n, v_0, \ldots, v_m]$. Notice $S(u_0, \ldots, u_n, v_0, \ldots, v_m)$ is a polynomial in two sets of variables $u_0, \ldots, u_n$ and $v_0, \ldots, v_m$. If $M = cu_0^{\alpha(0)} \cdots u_n^{\alpha(n)} v_0^{\beta(0)} \cdots v_m^{\beta(m)}$ is a nonzero monomial in

$$\mathbb{Z}[u_0, \ldots, u_n, v_0, \ldots, v_m]$$
$$(c \in \mathbb{Z}^*, \text{ and } (\alpha(0), \ldots, \alpha(n), \beta(0), \ldots, \beta(m)) \in \mathbb{N}_0^{m+n+2})$$

then clearly $\partial(M) = \sum_{i=0}^{n} \alpha(i) + \sum_{j=0}^{m} \beta(j)$. Each nonzero monomial M also has a weight which is defined as follows:

Definition 8.25 Let $M = cu_0^{\alpha(0)} \cdots u_n^{\alpha(n)} v_0^{\beta(0)} \cdots v_m^{\beta(m)}$ be a nonzero monomial in $\mathbb{Z}[u_0, \ldots, u_n, v_0, \ldots, v_m]$. The weight of M is the integer $\sum_{i=0}^{n} i\alpha(i) + \sum_{j=0}^{m} j\beta(j)$.

We will let $\omega(M)$ denote the weight of M. For example,

$$\omega(u_0^2 v_1^3) = 0(2) + 1(3) = 3$$

and

$$\omega(u_0^2 u_1^3 u_2^2 v_0^5 v_1^2 v_3^2) = 0(2) + 1(3) + 2(2) + 0(5) + 1(2) + 3(2) = 15$$

Any nonzero $f \in \mathbb{Z}[u_0, \ldots, u_n, v_0, \ldots, v_m]$ can be written uniquely in the form $r_1 M_1 + \cdots + r_t M_t$ where $r_1, \ldots, r_t \in \mathbb{Z}^*$ and $M_1, \ldots, M_t$ are distinct monic monomials in $u_0, \ldots, u_n, v_0, \ldots, v_m$.

Definition 8.26 Let $f = r_1 M_1 + \cdots + r_t M_t$ be a nonzero polynomial as above. f is isobaric of weight p if $\omega(M_i) = p$ for all $i = 1, \ldots, t$.

Thus, if each monomial appearing in f has the same weight p, then f is called an isobaric polynomial of weight p. Consider the following examples.

Example 8.27 Let $f(u_0, u_1, v_0, v_1) = u_0^2 v_1^2 + 2u_0 u_1 v_0 v_1 - 3v_1^4$. Then f is a homogeneous polynomial of degree 4. The weight of each monomial in f is as follows: $\omega(u_0^2 v_1^2) = 2$, $\omega(2u_0 u_1 v_0 v_1) = 2$, and $\omega(-3v_1^4) = 4$. In particular, f is not isobaric.

Consider

$$S(u_0,u_1,u_2,v_0,v_1,v_2) = \det \begin{bmatrix} u_0 & u_1 & u_2 & 0 \\ 0 & u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 & 0 \\ 0 & v_0 & v_1 & v_2 \end{bmatrix}$$

$$= u_0^2 v_2^2 - u_0 u_1 v_1 v_2 + u_0 u_2 v_1^2 - u_0 u_2 v_0 v_2 + u_1^2 v_0 v_2 - u_0 u_2 v_0 v_2 - u_1 u_2 v_0 v_1 + u_2^2 v_0^2$$

Obviously $S(u_0,u_1,u_2,v_0,v_1,v_2)$ is a homogeneous polynomial of degree 4 and isobaric of weight 4. ∎

The second example of Example 8.27 is an illustration of the following general result.

Theorem 8.28 $S(u_0, \ldots, u_n, v_0, \ldots, v_m)$ is homogeneous of degree $m + n$ and isobaric of weight mn.

Proof. Let Y be an indeterminate over $\mathbb{Z}[u_0, \ldots, u_n, v_0, \ldots, v_m]$. Then

$$\begin{aligned} S(Yu_0, \ldots, Yu_n, Yv_0, \ldots, Yv_m) &= \det(Y\,\mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m)) \\ &= Y^{m+n} \det(\mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m)) \\ &= Y^{m+n} S(u_0, \ldots, u_n, v_0, \ldots, v_m) \end{aligned}$$

Thus, Lemma 8.5 implies $S(u_0, \ldots, u_n, v_0, \ldots, v_m)$ is a homogeneous polynomial of degree $m + n$.

Let $\mathcal{S} = \mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m)$ and $S = S(u_0, \ldots, u_n, v_0, \ldots, v_m)$. To show S is isobaric of weight mn, we need to examine each monomial appearing in S. From Definition 2.16, we see a typical monomial in $S = \det(\mathcal{S})$ has the following form:

$$\textbf{8.29} \quad \pm \prod_{i=1}^{m+n} [\mathcal{S}]_{i\sigma(i)} \quad \text{for some } \sigma \in S_{m+n}$$

Using equation 8.9, the product appearing in 8.29 can be written as follows:

$$\textbf{8.30} \quad \pm \prod_{i=1}^{m+n} [\mathcal{S}]_{i\sigma(i)} = \pm \left(\prod_{i=1}^{m} u_{\sigma(i)-i} \right) \left(\prod_{i=m+1}^{m+n} v_{m+\sigma(i)-i} \right)$$

If the monomial in equation 8.30 is nonzero, then its weight is

$$\begin{aligned} &\sum_{i=1}^{m} (\sigma(i) - i) + \sum_{i=m+1}^{m+n} (m + \sigma(i) - i) \\ &\qquad = \sum_{i=1}^{m+n} \sigma(i) - \sum_{i=1}^{m+n} i + \sum_{i=m+1}^{m+n} m = \sum_{i=m+1}^{m+n} m = mn \end{aligned}$$

Thus, $S(u_0, \ldots, u_n, v_0, \ldots, v_m)$ is isobaric of weight mn. ∎

We now focus our attention on resultants. Our first observation is $\mathcal{R}(f, g) \in (f, g)$; that is, the resultant of two polynomials f and g always lies in the ideal generated by f and g.

Theorem 8.31 Let

$$f(X) = a_0X^n + a_1X^{n-1} + \cdots a_{n-1}X + a_n$$

and

$$g(X) = b_0X^m + b_1X^{m-1} + \cdots + b_{m-1}X + b_m$$

be two polynomials in $R[X]$. We assume $a_0, b_0 \neq 0$ and $m + n \geq 1$. Then there exist $a(X)$ and $b(X)$ in $R[X]$ such that

(i) $\partial(a) \leq m - 1$ and $\partial(b) \leq n - 1$ and

(ii) $\mathcal{R}(f, g) = a(X)f(X) + b(X)g(X)$.

Proof. Let $\mathcal{S}(a_0, \ldots, a_n, b_0, \ldots, b_m)$ denote the $(m + n) \times (m + n)$ matrix formed by replacing $u_0, \ldots, u_n$ with $a_0, \ldots, a_n$ and $v_0, \ldots, v_m$ with $b_0, \ldots, b_m$, respectively. Then $\mathcal{R}(f, g) = \det(\mathcal{S}(a_0, \ldots, a_n, b_0, \ldots, b_m))$. A simple computation shows

$$\mathcal{S}(a_0, \ldots, a_n, b_0, \ldots, b_m) \begin{bmatrix} X^{n+m-1} \\ X^{n+m-2} \\ \vdots \\ X^n \\ X^{n-1} \\ \vdots \\ X \\ 1 \end{bmatrix} = \begin{bmatrix} X^{m-1}f(X) \\ X^{m-2}f(X) \\ \vdots \\ f(X) \\ X^{n-1}g(X) \\ \vdots \\ Xg(X) \\ g(X) \end{bmatrix} \tag{8.32}$$

Let $c_i = \text{cof}_{i,m+n}(\mathcal{S}(a_0, \ldots, a_n, b_0, \ldots, b_m))$ for $i = 1, \ldots, m + n$. Thus, $c_1, \ldots, c_{m+n}$ are the cofactors of the last column of $\mathcal{S}(a_0, \ldots, a_n, b_0, \ldots, b_m)$. From Laplace's expansion, we have

$$\sum_{i=1}^{m+n} c_i[\mathcal{S}(a_0, \ldots, a_n, b_0, \ldots, b_m)]_{i,m+n} = \mathcal{R}(f,g) \tag{8.33}$$

and

$$\sum_{i=1}^{m+n} c_i[\mathcal{S}(a_0, \ldots, a_n, b_0, \ldots, b_m)]_{i,j} = 0$$

when $j \neq m + n$.

Set $\mathcal{S} = \mathcal{S}(a_0, \ldots, a_n, b_0, \ldots, b_m)$. Multiplying row i of equation 8.32 with c_i and adding the results produce the following equations:

8.34
$$\begin{aligned} c_1X^{m-1}f(X) + c_2X^{m-2}f(X) + \cdots + c_mf(X) + c_{m+1}X^{n-1}g(X) \\ + c_{m+2}X^{n-2}g(X) + \cdots + c_{m+n}g(X) \\ = \sum_{j=1}^{m+n} c_1[\mathcal{S}]_{1j}X^{m+n-j} + \sum_{j=1}^{m+n} c_2[\mathcal{S}]_{2j}X^{m+n-j} + \cdots \\ + \sum_{j=1}^{m+n} c_{m+n}[\mathcal{S}]_{(m+n)j}X^{m+n-j} = \sum_{i=1}^{m+n}\sum_{j=1}^{m+n} c_i[\mathcal{S}]_{ij}X^{m+n-j} \\ = \sum_{j=1}^{m+n}\left(\sum_{i=1}^{m+n} c_i[\mathcal{S}]_{ij}\right)X^{m+n-j} = \mathcal{R}(f,g) \quad \text{(by 8.33)} \end{aligned}$$

Thus,

$$a(X)f(X) + b(X)g(X) = \mathcal{R}(f, g)$$

where

$$a(X) = c_1X^{m-1} + c_2X^{m-2} + \cdots + c_m$$

and

$$b(X) = c_{m+1}X^{n-1} + c_{m+2}X^{n-2} + \cdots + c_{m+n}.$$

Clearly, $\partial(a) \leq m - 1$ and $\partial(b) \leq n - 1$. This completes the proof of Theorem 8.31. ∎

Some of the most important applications of resultants occur in polynomial rings $R[X_1, \ldots, X_p]$ in more than one variable. Let $f(X_1, \ldots, X_p)$ be a nonzero polynomial in $R[X_1, \ldots, X_p]$. We can view $f(X_1, \ldots, X_p)$ as a polynomial in X_p with coefficients in the ring $R[X_1, \ldots, X_{p-1}]$. Thus,

$$\begin{aligned} f(X_1, \ldots, X_p) = a_0(X_1, \ldots, X_{p-1})X_p^d + a_1(X_1, \ldots, X_{p-1})X_p^{d-1} \\ + \cdots + a_{d-1}(X_1, \ldots, X_{p-1})X_p \\ + a_d(X_1, \ldots, X_{p-1}) \end{aligned}$$

Here

$$\begin{aligned} &a_0(X_1, \ldots, X_{p-1}), \ldots, a_d(X_1, \ldots, X_{p-1}) \\ &\quad \in R[X_1, \ldots, X_{p-1}], \qquad a_0(X_1, \ldots, X_{p-1}) \neq 0 \end{aligned}$$

and d is the degree of f as a polynomial in X_p. The integer d is called the degree of f with respect to X_p. We will use the notation $\partial_{X_p}(f)$ to denote the degree of

f with respect to the variable X_p. Obviously $\partial_{X_p}(f) \leq 0$ if and only if $f \in R[X_1, \ldots, X_{p-1}]$. There is of course nothing special about the variable X_p in this discussion. We can make similar remarks relative to any variable $X_1, \ldots, X_p$. In particular, $\partial_{X_i}(f)$ makes sense for any $i = 1, \ldots, p$.

Now suppose $f(X_1, \ldots, X_p)$ and $g(X_1, \ldots, X_p)$ are two nonzero polynomials in $R[X_1, \ldots, X_p]$. Select a variable, say X_i. If $\partial_{X_i}(f) = n$ and $\partial_{X_i}(g) = m$, then

$$f(X_1, \ldots, X_p) = a_0X_i^n + a_1X_i^{n-1} + \cdots + a_{n-1}X_i + a_n$$

and

$$g(X_1, \ldots, X_p) = b_0X_i^m + b_1X_i^{m-1} + \cdots + b_{m-1}X_i + b_m$$

Here

$$a_0, \ldots, a_n, b_0, \ldots, b_m \in R[X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p]$$

and $a_0 \neq 0 \neq b_0$. If we assume $m + n \geq 1$, then we can form the determinant $S(a_0, \ldots, a_n, b_0, \ldots, b_m)$. We will call this determinant the resultant of f and g with respect to X_i. In this book, we will let $\mathcal{R}_{X_i}(f, g)$ denote the resultant of f and g with respect to X_i. Thus,

$$\mathcal{R}_{X_i}(f, g) = S(a_0, \ldots, a_n, b_0, \ldots, b_m) \in R[X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p]$$

Consider the following simple example of two variables.

Example 8.35 Let

$$\begin{aligned} f(X, Y) &= X^2 + Y^2 + XY + X + 1 \\ g(X, Y) &= XY + X + Y + 1 \end{aligned} \qquad \text{in } \mathbb{Z}[X, Y]$$

To compute $\mathcal{R}_X(f, g)$, write f and g as polynomials in X with coefficients in $\mathbb{Z}[Y]$.

8.36
$$\begin{aligned} f(X, Y) &= X^2 + (1 + Y)X + (1 + Y^2) && [n = 2] \\ g(X, Y) &= (1 + Y)X + (1 + Y) && [m = 1] \end{aligned}$$

Therefore,

8.37
$$\mathcal{R}_X(f,g) = \det \begin{bmatrix} 1 & 1 + Y & 1 + Y^2 \\ 1 + Y & 1 + Y & 0 \\ 0 & 1 + Y & 1 + Y \end{bmatrix} = (1 + Y)^2 (1 - Y + Y^2)$$

To compute $\mathcal{R}_Y(f, g)$, write f and g as polynomials in Y with coefficients in $\mathbb{Z}[X]$.

8.38 $f(X, Y) = Y^2 + XY + (X^2 + X + 1) \quad [n = 2]$
$g(X, Y) = (1 + X)Y + (1 + X) \quad [m = 1]$

Therefore,

8.39 $$\mathcal{R}_Y(f,g) = \det \begin{bmatrix} 1 & X & X^2 + X + 1 \\ 1 + X & 1 + X & 0 \\ 0 & 1 + X & 1 + X \end{bmatrix} = (1 + X)^2 (2 + X^2)$$ ∎

Again suppose f and g are nonzero polynomials in $R[X_1, \ldots, X_p]$. Suppose $\partial_{X_p}(f) = n$ and $\partial_{X_p}(g) = m$ with $m + n \geq 1$. Then we have

8.40 $f = a_0 X_p^n + a_1 X_p^{n-1} + \cdots + a_{n-1} X_p + a_n$
$g = b_0 X_p^m + b_1 X_p^{m-1} + \cdots + b_{m-1} X_p + b_m$

In equation 8.40, $a_0, \ldots, a_n, b_0, \ldots, b_m \in R[X_1, \ldots, X_{p-1}]$, and $a_0 \neq 0 \neq b_0$. Suppose we also assume $f(X_1, \ldots, X_p)$ is homogeneous of degree n and $g(X_1, \ldots, X_p)$ is homogeneous of degree m. Then each polynomial $a_i(X_1, \ldots, X_{p-1})$ appearing in equation 8.40 is either zero or a homogeneous polynomial (in $R[X_1, \ldots, X_{p-1}]$) of degree i. Similarly, each $b_j(X_1, \ldots, X_{p-1})$ is either zero or a homogeneous polynomial of degree j. We claim $\mathcal{R}_{X_p}(f, g)$ is either zero or a homogeneous polynomial of degree mn in $R[X_1, \ldots, X_{p-1}]$. (See Exercise 20 for a proof using Theorem 8.28.)

To see this, set

$$h(X_1, \ldots, X_{p-1}) = \mathcal{R}_{X_p}(f, g) = S(a_0, \ldots, a_n, b_0, \ldots, b_m)$$

Let Y be an indeterminate over $R[X_1, \ldots, X_{p-1}]$. Since each a_i (or b_j) is a homogeneous polynomial of degree i (or j), we have

8.41 $h(YX_1, \ldots, YX_{p-1})$

$$= \det \begin{bmatrix} a_0 & Ya_1 & Y^2 a_2 & \cdots & Y^n a_n & 0 & 0 & \cdots & 0 \\ 0 & a_0 & Ya_i & \cdots & Y^{n-1} a_{n-1} & Y^n a_n & 0 & \cdots & 0 \\ \vdots & \vdots & & & \vdots & & & & \vdots \\ \vdots & \vdots & & & \vdots & & & & \vdots \\ \vdots & \vdots & & & \vdots & & & & \vdots \\ 0 & 0 & \cdots & a_0 & Ya_1 \cdots & & & & Y^n a_n \\ b_0 & Yb_1 & \cdots & & Y^m b_m & 0 & \cdots & & 0 \\ 0 & b_0 & \cdots & & Y^{m-1} b_{m-1} & Y^m b_m & \cdots & & 0 \\ \vdots & \vdots & & & \vdots & & & & \vdots \\ \vdots & \vdots & & & \vdots & & & & \vdots \\ \vdots & \vdots & & & \vdots & & & & \vdots \\ 0 & 0 & & b_0 & Yb_1 \cdots & & & & Y^m b_m \end{bmatrix}$$

Let us denote the $(m + n) \times (m + n)$ matrix whose determinant appears in equation 8.41 by the letter C. Let L denote the $(m + n) \times (m + n)$ matrix obtained from C by multiplying the first m rows of C with $Y, Y^2, \ldots, Y^m$, respectively, and the last n rows of C by $Y, \ldots, Y^n$, respectively. Thus, L has the following form:

8.42
$$L = \begin{bmatrix} a_0Y & a_1Y^2 & a_nY^{n+1} & 0 & \ldots & 0 \\ 0 & a_0Y^2 & a_{n-1}Y^{n+1} & a_nY^{n+2} & 0\ldots & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & a_0Y^m \quad a_1Y^{m+1} & \ldots & & a_nY^{m+n} \\ b_0Y & b_1Y^2 & b_mY^{m+1} & 0 & \ldots & 0 \\ 0 & b_0Y^2 & b_{m-1}Y^{m+1} & b_mY^{m+2} & 0\ldots & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & b_0Y^n & & & b_mY^{m+n} \end{bmatrix}$$

Using equation 8.9, we have

8.43
$$[L]_{ij} = Y^j a_{j-i} \quad \text{for } i = 1, \ldots, m; j = 1, \ldots, m + n$$
$$[L]_{m+ij} = Y^j b_{j-i} \text{ for } i = 1, \ldots, n; j = 1, \ldots, m + n$$

In equation 8.43, we use the same conventions as in equation 8.9: $a_k = 0$ if $k < 0$ or $k > n$ and $b_k = 0$ if $k < 0$ or $k > m$. Notice that Y^j is the power of Y which appears in each $[L]_{ij}$ as i varies from 1 to $m + n$.

Since multiplying a row of a determinant by a constant multiplies the determinant by that constant, we have

8.44 $\det(L) = Y^\alpha h(YX_1, \ldots, YX_{p-1})$

In equation 8.44,

$$\alpha = (1 + 2 + \cdots + n) + (1 + 2 + \cdots m)$$
$$= \frac{m(m + 1)}{2} + \frac{n(n + 1)}{2}$$

On the other hand, we can pull Y^j out of each column of $\det(L)$ as j varies from 1 to $m + n$. Therefore,

8.45 $\det(L) = Y^\beta h(X_1, \ldots, X_{p-1})$

Here $\beta = 1 + 2 + \cdots + (m + n) = (m + n)(m + n + 1)/2$. Equations 8.44 and 8.45 imply

8.46 $h(YX_1, \ldots, YX_{p-1}) = Y^{\beta-\alpha}h(X_1, \ldots, X_{p-1})$

Since $\beta - \alpha = mn$, we conclude $h(X_1, \ldots, X_{p-1}) = \mathcal{R}_{X_p}(f, g)$ is homogeneous of degree mn (or zero).

The argument just given works for any variable as long as equations like those in equation 8.40 are available. Hence, we have proved the following lemma.

Lemma 8.47 Let $f(X_1, \ldots, X_p)$ and $g(X_1, \ldots, X_p)$ be homogeneous polynomials of degrees n and m, respectively, in $R[X_1, \ldots, X_p]$. Suppose for some variable X_i, we have

$$f = a_0X_i^n + a_1X_i^{n-1} + \cdots + a_{n-1}X_i + a_n$$

and

$$g = b_0X_i^m + b_1X_i^{m-1} + \cdots + b_{m-1}X_i + b_m$$

Here

$$a_0, \ldots, a_n, b_0, \ldots, b_m \in R[X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p]$$

and $a_0 \neq 0 \neq b_0$. We assume $m + n \geq 1$. Then $\mathcal{R}_{X_i}(f, g)$ is either zero or a homogeneous polynomial of degree mn in $R[X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p]$. ∎

Consider the following examples.

Example of 8.48 Let $f, g \in \mathbb{Z}[X_1, X_2, X_3, X_4]$.

(a) If

$$f = 2X_1^2 + X_1X_2 \qquad [n = 2, a_0 = 2, a_1 = X_2, a_2 = 0]$$

and

$$g = X_1 - X_3 \qquad [m = 1, b_0 = 1, b_1 = -X_3]$$

then f and g are homogeneous polynomials of degrees 2 and 1, respectively, which satisfy the hypotheses of Lemma 8.47 with respect to X_1. Then

$$\mathcal{R}_{X_1}(f,g) = \det \begin{bmatrix} 2 & X_2 & 0 \\ 1 & -X_3 & 0 \\ 0 & 1 & -X_3 \end{bmatrix} = 2X_3^2 + X_2X_3$$

a homogeneous polynomial of degree 2 in $\mathbb{Z}[X_2, X_3, X_4]$.

(b) If

$$f = 2X_1^2 + X_1X_2 \ [n = 2, a_0 = 2, a_1 = X_2, a_2 = 0]$$

and

$$g = X_1^2 + X_3X_4 \ [m = 2, b_0 = 1, b_1 = 0, b_2 = X_3X_4]$$

then f and g are homogeneous polynomials of degree 2 which satisfy the hypotheses of Lemma 8.47 with respect to X_1. Then

$$\mathcal{R}_{X_1}(f,g) = \det \begin{bmatrix} 2 & X_2 & 0 & 0 \\ 0 & 2 & X_2 & 0 \\ 1 & 0 & X_3X_4 & 0 \\ 0 & 1 & 0 & X_3X_4 \end{bmatrix} = 4X_3^2X_4^2 + X_2^2X_3X_4$$

a homogeneous polynomial of degree 4 in $\mathbb{Z}[X_2,X_3,X_4]$. ∎

We should point out that the hypotheses in Lemma 8.47 are not always satisfied. In general, $\mathcal{R}_{X_i}(f, g)$ may not have degree mn. Consider the following simple example.

Example 8.49 Let $f = X_1X_2$ and $g = X_3X_4$ in $\mathbb{Z}[X_1, X_2, X_3, X_4]$. Again both f and g are homogeneous polynomials of degree 2. However, $\partial_{X_1}(f) = 1$, $\partial_{X_2}(f) = 1$, $\partial_{X_3}(f) = 0$, and $\partial_{X_4}(f) = 0$. In particular, f cannot be written as in Lemma 8.47 for an choice of variable X_1, X_2, X_3, or X_4. Similarly, g cannot be written as in Lemma 8.47 either.

Suppose we view f and g as polynomials in X_1. Then $\mathcal{R}_{X_1}(f, g) = X_3X_4$, a homogeneous polynomial of degree 2 (not 4) in $\mathbb{Z}[X_2, X_3, X_4]$. The point here is that if the hypotheses of Lemma 8.47 are not satisfied, then the degree of $\mathcal{R}_{X_i}(f, g)$ need not be mn. ∎

We can now derive the following useful formula for resultants.

Theorem 8.50 Let R be an integral domain. Let

$$f(X) = a_0X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n$$

and

$$g(X) = b_0X^m + b_1X^{m-1} + \cdots + b_{m-1}X + b_m$$

be two polynomials in $R[X]$ such that $a_0 \neq 0 \neq b_0$ and $m,n \geq 1$. Let $\{\xi_1, \ldots, \xi_n\}$ and $\{\theta_1, \ldots, \theta_m\}$ be the roots of f and g, respectively, in some splitting field E of $f(X)g(X)$ over $Q(R)$. Then $\mathcal{R}(f, g) = a_0^m \prod_{i=1}^n g(\xi_i)$ and $\mathcal{R}(f, g) = (-1)^{mn}b_0^n \prod_{j=1}^m f(\theta_j)$.

Proof. Let $K = Q(R)$, the quotient field of R. If the theorem is true for K, it is certainly true for R. Hence, we can assume $R = K$ a field. Since $a_0 \neq 0 \neq b_0$, $f(X) = a_0 f_1(X)$ and $g(X) = b_0 g_1(X)$ for monic polynomials $f_1, g_1 \in K[X]$. Furthermore, the roots of $f_1(X)$ (and $g_1(X)$) in E are exactly the same as those of $f(X)$ $(g(X))$. Suppose the theorem is true for monic polynomials in $K[X]$. Then

$$\begin{aligned}\mathcal{R}(f, g) &= \mathcal{R}(a_0 f_1(X), b_0 g_1(X)) = a_0^m b_0^n \mathcal{R}(f_1, g_1) = a_0^m b_0^n \prod_{i=1}^{n} g_1(\xi_i) \\ &= a_0^m \prod_{i=1}^{n} b_0 g_1(\xi_i) = a_0^m \prod_{i=1}^{m} g(\xi_i)\end{aligned}$$

Similarly, $\mathcal{R}(f, g) = (-1)^{mn} b_0^n \prod_{j=1}^{m} f(\theta_j)$. Hence, we can assume that f and g are both monic polynomials $(a_0 = 1 = b_0)$ in $K[X]$.

Since f and g are both monic polynomials, $f(X) = \prod_{i=1}^{n}(X - \xi_i)$ and $g(X) = \prod_{j=1}^{m}(X - \theta_j)$ in $E[X]$.

Let $x_1, \ldots, x_n, y_1, \ldots, y_m$ and X be indeterminates over $\mathbb{Q}$. Define the following polynomials in $\mathbb{Z}[X, x_1, \ldots, x_n, y_1, \ldots, y_m]$.

$$\begin{aligned}\textbf{8.51}\quad F(X, x_1, \ldots, x_n, y_1, \ldots, y_m) &= \prod_{i=1}^{n} (X - x_i) \\ G(X, x_1, \ldots, x_n, y_1, \ldots, y_m) &= \prod_{j=1}^{m} (X - y_j) \\ H(x_1, \ldots, x_n, y_1, \ldots, y_m) &= \prod_{i=1}^{n} \prod_{j=1}^{m} (x_i - y_j)\end{aligned}$$

Clearly, F, G, and H are homogeneous polynomials of degress n, m, and mn, respectively, in $\mathbb{Z}[X, x_1, \ldots, x_n, y_1, \ldots, y_m]$. Set $P(x_1, \ldots, x_n, y_1, \ldots, y_m) = \mathcal{R}_X(F, G)$.

It is easy to see that F and G satisfy the hypotheses of Lemma 8.47 with respect to X. Hence, $P(x_1, \ldots, x_n, y_1, \ldots, y_m)$ is either zero or a homogeneous polynomial of degree mn in $\mathbb{Z}[x_1, \ldots, x_n, y_1, \ldots, y_m]$. Clearly, $P(x_1, \ldots, x_n, y_1, \ldots, y_m) \neq 0$. For we can always choose a specialization of $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ in $\mathbb{Q}$ so that the corresponding $\overline{F}$ and $\overline{G}$ have no common roots. Then $\mathcal{R}(\overline{F}, \overline{G}) \neq 0$ by Corollary 8.23. Since $\mathcal{R}(\overline{F}, \overline{G})$ is obviously a specialization of $\mathcal{R}_X(F, G)$, we conclude $\mathcal{R}_X(F, G) \neq 0$. Thus, $P(x_1, \ldots, x_n, y_1, \ldots, y_m)$ is a nonzero, homogeneous polynomial of degree mn in $\mathbb{Z}[x_1, \ldots, x_n, y_1, \ldots, y_m]$.

$H(x_1, \ldots, x_n, y_1, \ldots, y_m)$ is also a nonzero, homogeneous polynomial of degree mn in $\mathbb{Z}[x_1, \ldots, x_n, y_1, \ldots, y_m]$. We claim $H \mid P$. To see this, fix $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$ and consider the monomial $x_i - y_j$. If we replace x_i with y_j in F in equation 8.51, then F and G have a common factor $X - y_j$. Therefore, Corollary 8.22 implies

$$P(x_1, \ldots, x_{i-1}, y_j, x_{i+1}, \ldots, x_n, y_1, \ldots, y_m) = 0.$$

We can think of $P(x_1, \ldots, x_n, y_1, \ldots, y_m)$ as a polynomial in

$$(\mathbb{Z}[x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n, y_1, \ldots, y_m])[x_i].$$

Then $P(x_1, \ldots, x_{i-1}, y_j, x_{i+1}, \ldots, x_n, y_1, \ldots, y_m) = 0$ means $x_i - y_j$ divides $P(x_1, \ldots, x_n, y_1, \ldots, y_m)$ in

$$\begin{aligned}&\mathbb{Z}[x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n, y_1, \ldots, y_m][x_i]\\ &\quad = \mathbb{Z}[x_1, \ldots, x_n, y_1, \ldots, y_m] \quad \text{(Corollary 7.11)}\end{aligned}$$

Since i and j are arbitrary, we conclude $x_i - y_j \mid P(x_1, \ldots, x_n, y_1, \ldots, y_m)$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Since these monomials $x_i - y_j$ ($1 \le i \le n$, $1 \le j \le m$) are all nonassociate, irreducible polynomials, it follows that $H \mid P$ in $\mathbb{Z}[x_1, \ldots, x_n, y_1, \ldots, y_m]$ as claimed.

We have already noted that H and P are nonzero, homogeneous polynomials of degree mn in $\mathbb{Z}[x_1, \ldots, x_n, y_1, \ldots, y_m]$. Since $H \mid P$, we must have $P(x_1, \ldots, x_n, y_1, \ldots, y_m) = cH(x_1, \ldots, x_n, y_1, \ldots, y_m)$ for some nonzero integer c. The reader will notice that c is derived from the definitions in equation 8.51. In particular, the value of c does not depend on $f(X)$ or $g(X)$. In fact, the nonzero integer c is defined by the following equation in $\mathbb{Z}[x_1, \ldots, x_n, y_1, \ldots, y_m]$:

8.52 $$\begin{aligned}&\mathcal{R}_X(F(X,x_1, \ldots, x_n,y_1, \ldots, y_m), G(X,x_1, \ldots, x_n,y_1, \ldots, y_m))\\ &\quad = c \prod_{i=1}^{n} \prod_{j=1}^{m} (x_i - y_j)\end{aligned}$$

We now specialize equation 8.52 by replacing $x_1, \ldots, x_n$ with $\xi_1, \ldots, \xi_n$ and $y_1, \ldots, y_m$ with $\theta_1, \ldots, \theta_m$, respectively. The polynomial $F(X, x_1, \ldots, x_n, y_1, \ldots, y_m)$ specializes to $\prod_{i=1}^{n} (X - \xi_i) = f(X)$. $G(X, x_1, \ldots, x_n, y_1, \ldots, y_m)$ specializes to $\prod_{j=1}^{m} (X - \theta_j) = g(X)$. Hence, the equation in 8.52 becomes

8.53 $$\mathcal{R}(f,g) = c \prod_{i=1}^{n} \prod_{j=1}^{m} (\xi_i - \theta_j)$$

In particular,

$$\mathcal{R}(f,g) = cg(\xi_1) \cdot \cdots g(\xi_n) = c \prod_{i=1}^{n} g(\xi_i)$$

We also get

$$\begin{aligned}\mathcal{R}(f,g) &= c \prod_{j=1}^{m} \prod_{i=1}^{n} -(\theta_j - \xi_i) = c \prod_{j=1}^{m} (-1)^n f(\theta_j)\\ &= c(-1)^{mn} \prod_{j=1}^{m} f(\theta_j)\end{aligned}$$

Thus, the proof of the theorem will be complete when we show $c = 1$.

We evaluate c by a little trickery. Suppose $X, z_1, \ldots, z_m$ are indeterminates over $\mathbb{Z}$. Let $f(X) = (X - 1)^n$ and $g(X) = X^m + z_1X^{m-1} + \cdots + z_{m-1}X + z_m$. Then $f(X)$ and $g(X)$ are polynomials in $(\mathbb{Z}[z_1, \ldots, z_m])[X]$. Since $R = \mathbb{Z}[z_1, \ldots, z_m]$ is an integral domain, we can apply the above proof to $f(X)$ and $g(X)$. Keep in mind that c is defined by equation 8.52 and consequently does not depend on f and g.

Write

$$f(X) = X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n \in \mathbb{Z}[X].$$

Our comments about the diagonal of Sylvester's matrix (after equation 8.9) readily imply the coefficient of z_m^n in $\mathcal{R}(f, g) = S(1, a_1, \ldots, a_n, 1, z_1, \ldots, z_m)$ is 1. The above proof applied to f and g gives us

$$\mathcal{R}(f,g) = c \prod_{i=1}^{n} g(1) = c \prod_{i=1}^{n}(1 + z_1 + \cdots + z_m)$$

The coefficient of z_m^n in this last product is c. Therefore, $c = 1$, and the proof of Theorem 8.50 is complete. ∎

There is one important application of Theorem 8.50 which we will use in the next chapter.

Theorem 8.54 Let $A \in M_{n\times n}(R)$, and let $g(X) \in R[X]^*$. Then $\mathcal{R}(C_A(X), g(X)) = \det(g(A))$.

Proof. Suppose $\partial(g) = 0$. Then $g(X)$ is a nonzero constant $b \in R$. Then $\mathcal{R}(C_A(X), b) = b^n$. On the other hand, $g(A) = bI_n$ and $\det(g(A)) = \det(bI_n) = b^n$. Thus, the theorem is proved if $\partial(g) = 0$. Hence, we can assume $\partial(g) \geq 1$.

We first prove the theorem when $R = \mathbb{Z}$. Suppose

$$\textbf{8.55}\quad C_A(X) = \prod_{i=1}^{n} (X - \xi_i) \quad \text{and} \quad g(X) = b \prod_{j=1}^{m} (X - \theta_j)$$

in some splitting field E of $C_A(X)g(X)$ over $\mathbb{Q}$. In equation 8.55, b $(\in \mathbb{Z}^*)$ is the leading coefficient of $g(X)$ in $E[X]$. By Theorem 8.50,

$$\textbf{8.56}\quad \mathcal{R}(C_A(X),g(X)) = \prod_{i=1}^{n} g(\xi_i)$$

On the other hand, in $M_{n\times n}(E)$, we have

$$g(A) = b \prod_{j=1}^{m} (A - \theta_jI_n) = b(-1)^m \prod_{j=1}^{m} (\theta_jI_n - A)$$

Therefore,

$$\begin{aligned}\det(g(A)) &= b^n(-1)^{mn}\prod_{j=1}^{m}\det(\theta_j I_n - A) = b^n(-1)^{mn}\prod_{j=1}^{m}C_A(\theta_j)\\ &= b^n(-1)^{mn}\prod_{j=1}^{m}\prod_{i=1}^{n}(\theta_j - \xi_i) = \prod_{i=1}^{n}\left(b\prod_{j=1}^{m}(\xi_i - \theta_j)\right)\\ &= \prod_{i=1}^{n}g(\xi_i) = \mathcal{R}(C_A(X), g(X))\end{aligned}$$

Thus, Theorem 8.54 is proved when $R = \mathbb{Z}$.

To prove the theorem in general, we will use the Specialization Lemma 8.6. Suppose R is an arbitrary commutative ring. Let $A \in M_{n\times n}(R)$ and $g(X) \in R[X]^*$. We have already observed we can assume $\partial(g) = m \geq 1$. Suppose $A = (a_{ij})$ and $g(X) = b_0X^m + b_1X^{m-1} + \cdots + b_{m-1}X + b_m$. Here $b_0 \neq 0$.

Let $\{x_{ij} \mid 1 \leq i, j \leq n\}$, $\{y_0, \ldots, y_m\}$, and X all be independent indeterminates over $\mathbb{Z}$. Set $\overline{R} = \mathbb{Z}[x_{11}, \ldots, x_{nn}, y_0, \ldots, y_m]$, $\overline{A} = (x_{ij}) \in M_{n\times n}(\overline{R})$, and $\overline{g}(X) = y_0X^m + y_1X^{m-1} + \cdots + y_{m-1}X + y_m \in \overline{R}[X]$. Define two polynomials F and G in $\overline{R}$ as follows:

8.57
$$\begin{aligned}F(x_{11}, \ldots, x_{nn}, y_0, \ldots, y_m) &= \mathcal{R}_X(C_{\overline{A}}(X), \overline{g}(X))\\ G(x_{11}, \ldots, x_{nn}, y_0, \ldots, y_m) &= \det(\overline{g}(\overline{A}))\end{aligned}$$

The first part of this proof (when $R = \mathbb{Z}$) implies

$$F(j_{11}, \ldots, j_{nn}, j_0, \ldots, j_m) = G(j_{11}, \ldots, j_{nn}, j_0, \ldots, j_m)$$

for all positive integers $j_{11}, \ldots, j_{nn}, j_0, \ldots, j_n$. Hence, the Specialization Lemma implies

$$F(a_{11}, \ldots, a_{nn}, b_0, \ldots, b_m) = G(a_{11}, \ldots, a_{nn}, b_0, \ldots, b_m)$$

In particular, $\mathcal{R}(C_A(X), g(X)) = \det(g(A))$. ∎

We finish this chapter with an example of Theorem 8.54.

Example 8.58 (a) Let $R = \mathbb{Z}$ and $g(X) = 3X + 2 \in \mathbb{Z}[X]$. Set

$$A = \begin{bmatrix}1 & 2\\ 2 & 3\end{bmatrix} \in M_{2\times 2}(\mathbb{Z})$$

Then $C_A(X) = X^2 - 4X - 1$. Therefore,

$$\mathcal{R}(C_A(X), g(X)) = \det\begin{bmatrix}1 & -4 & -1\\ 3 & 2 & 0\\ 0 & 3 & 2\end{bmatrix} = 19$$

On the other hand,

$$g(A) = 3A + 2I_2 = \begin{bmatrix} 5 & 6 \\ 6 & 11 \end{bmatrix}$$

Therefore, $\det(g(A)) = 19 = \mathcal{R}(C_A(X), g(X))$.
(b) Suppose $R = \mathbb{Z}/6\mathbb{Z} = \{0, 1, 2, 3, 4, 5\}$. Again let $g(X) = 3X + 2 \in R[X]$ and

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \in M_{2\times 2}(R)$$

Then $C_A(X) = X^2 + 2X + 5$, and

$$\mathcal{R}(C_A(X), g(x)) = \det \begin{bmatrix} 1 & 2 & 5 \\ 3 & 2 & 0 \\ 0 & 3 & 2 \end{bmatrix} = 1$$

On the other hand

$$g(A) = 3A + 2I_2 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$$

Therefore, $\det(g(A)) = 1 = \mathcal{R}(C_A(X), g(X))$. ∎

There are several applications of resultants such as classical elimination theory, Hilbert's Nullstellensatz, and Bezout's theorem which are standard fare for a course on resultants. We have included this material in Appendix C at the end of this text.

EXERCISES

1. Let $f(X_1, \ldots, X_p)$ and $g(X_1, \ldots, X_p)$ be two homogeneous polynomials of degrees n and m, respectively, in $R[X_1, \ldots, X_p]$. Show fg is either zero or homogeneous of degree mn.
2. Let R be an integral domain. Suppose $f(X_1, \ldots, X_p) = g(X_1, \ldots, X_p)h(X_1, \ldots, X_p)$ in $R[X_1, \ldots, X_p]$. If f is homogeneous, show g and h are also homogeneous polynomials.
3. Use the results of Exercise 2 to show $X_1X_4 - X_2X_3$ is irreducible in $\mathbb{Z}[X_1, X_2, X_3, X_4]$.
4. Is Exercise 2 true if R is not a domain?

5. Let

$$V = \begin{bmatrix} 1 & X_1 & X_1^2 & \cdots & X_1^{n-1} \\ 1 & X_2 & X_2^2 & \cdots & X_2^{n-1} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & X_n & X_n^2 & \cdots & X_n^{n-1} \end{bmatrix} \in M_{n\times n}(\mathbb{Z}[X_1, \ldots, X_n])$$

The matrix V is called the Vandermonde matrix. Show $\det(V^tV)$ is a homogeneous polynomial of degree $n(n-1)$ in $\mathbb{Z}[X_1, \ldots, X_n]$.

6. Let $X_0, \ldots, X_p$ be indeterminates over R. Let $f(X_1, \ldots, X_p) \in R[X_1, \ldots, X_p]$. Set ${}^hf = X_0^{\partial(f)} f(X_1/X_0, \ldots, X_p/X_0)$. Prove the following assertions about the map ${}^h(*) : R[X_1, \ldots, X_p] \mapsto R[X_0, \ldots, X_p]$:
   (a) ${}^hf$ is a homogeneous polynomial of degree $\partial(f)$ in $R[X_0, \ldots, X_p]$.
   (b) ${}^h(fg) = {}^hf\,{}^hg$.
   (c) $X_0^{\partial(f)+\partial(g)}\,{}^h(f+g) = X_0^{\partial(f+g)}[X_0^{\partial(g)}\,{}^hf + X_0^{\partial(f)}\,{}^hg]$.
7. Let S be a commutative R-algebra. Let $s_1, \ldots, s_n \in S$. Verify that the map $\psi(s_1, \ldots, s_n) : R[X_1, \ldots, X_n] \mapsto S$ is an R-algebra homomorphism whose image is the R-subalgebra of S generated by $s_1, \ldots, s_n$.
8. Let $F(X_1, \ldots, X_n) \in \mathbb{Z}[X_1, \ldots, X_n]$. Let Λ be an infinite subset of $\mathbb{Z}$. Show $F(X_1, \ldots, X_n) = 0$ if and only if $F(j_1, \ldots, j_n) = 0$ for all $(j_1, \ldots, j_n) \in \Lambda^n$.
9. Verify the following remark made in the proof of Corollary 8.21: We can assume with no loss of generality that f and g are monic polynomials in $F[X]$.
10. Suppose R is a unique factorization domain. Let $f(X)$ and $g(X)$ be two polynomials of positive degree in $R[X]$. Show f and g have a common factor of positive degree if and only if $q(X)f(X) = p(X)g(X)$ for two nonzero polynomials q and p in $R[X]$ with $\partial(p) \leq \partial(f)$ and $\partial(q) \leq \partial(g)$.
11. Use the results in Exercise 10 to give a proof of Corollary 8.22.
12. The special cases $n \geq 1$, $m = 0$ and $n = 0$, $m \geq 1$ are of some interest in Theorem 8.31. Write out what the theorem says in these cases and verify the results directly.
13. Let R be a unique factorization domain of characteristic zero (i.e., $\mathbb{Z} \subseteq R$). Let $f(X) \in R[X]$ be a polynomial of positive degree. The resultant $\mathcal{R}(f, f')$ of f and f' (the derivative of f) is called the discriminant of f. Compute the discriminant of the following polynomials:
   (a) $f(X) = aX^2 + bX + c$.
   (b) $f(X) = aX^3 + bX^2 + cX + d$.

14. In Exercise 13, suppose $\partial(f) \geq 2$. Show there exists an irreducible $g(X) \in R[X]$ such that $\partial(g) \geq 1$ and $g^2 \mid f$ if and only if $\mathcal{R}(f, f') = 0$.
15. Compute $\mathcal{R}_{X_1}(f, g)$ and $\mathcal{R}_{X_2}(f, g)$ for the following two polynomials: $f(X_1, X_2) = X_1^2X_2 + X_1X_2 - 2X_1 + 1$, $g(X_1, X_2) = X_1X_2 + 2X_2 + 3$.
16. Find the points of intersection of the circle $x^2 + y^2 + 4x - 2y + 3 = 0$ and the hyperbola $x^2 + 4xy - y^2 + 10y - 9 = 0$. Use Corollary 8.21 for this problem.
17. Find all a and b such that

$$X^3 - 6X^2 + aX - 3 = 0$$
$$X^3 - X^2 + bX + 2 = 0$$

have two common solutions.

18. Find all solutions $(z_1, z_2)^t \in \mathbb{C}^2$ to the equations

$$X_1^3 + 2X_1^2X_2 + 2X_2(X_2 - 2)X_1 + X_2^2 - 4 = 0$$
$$X_1^2 + 2X_1X_2 + 2X_2^2 - 5X_2 + 2 = 0$$

Hint: Compute the resultant with respect to X_1 and use Corollary 8.21.

19. Let $f(X)$, $g(X)$, $h(X)$ be polynomials of positive degree in $\mathbb{Z}[X]$. Show $\mathcal{R}(f, gh) = \mathcal{R}(f, g)\mathcal{R}(f, h)$. Hint: Use Theorem 8.50.
20. Give another proof of Lemma 8.47 using Theorem 8.28 and a judicious interpretation of weight.

9
Zero Divisors in $M_{n\times n}(R)$

Let R be a commutative ring, and set $T = M_{n\times n}(R)$. A matrix $A \in T$ is called a left zero divisor in T if $AB = O$ for some nonzero matrix $B \in T$. Similarly, A is called a right zero divisor in T if $CA = O$ for some nonzero $C \in T$. In $M_{n\times n}(R)$, a matrix is a left zero divisor if and only if it is a right zero divisor. This follows from our first theorem in this section.

Theorem 9.1 Let $A \in M_{n\times n}(R)$.

(a) A is a left zero divisor in $M_{n\times n}(R)$ if and only if $\det(A) \in Z(R)$.

(b) A is a right zero divisor in $M_{n\times n}(R)$ if and only if $\det(A) \in Z(R)$.

Proof. Suppose $\det(A) \in Z(R)$. Then 4.11e implies $\mathrm{rk}(A) < n$. Therefore, Theorem 5.3 implies $A\xi = O$ for some nonzero vector $\xi \in R^n$. Set $B = (\xi \mid \xi \mid \cdots \mid \xi) \in M_{n\times n}(R)$. Then $B \neq O$, and $AB = (A\xi \mid \cdots \mid A\xi) = (O \mid \cdots \mid O) = O$. Thus, A is a left zero divisor in $M_{n\times n}(R)$.

Since $\det(A^t) = \det(A)$, the same proof shows $\det(A) \in Z(R)$ implies A^t is a left zero divisor in $M_{n\times n}(R)$. Hence, $A^tB = O$ for some nonzero $B \in M_{n\times n}(R)$. Then $B^t \neq O$, and $(B^tA)^t = A^tB = O$. Therefore, $B^tA = O$, and A is a right zero divisor in $M_{n\times n}(R)$.

Hence, if $\det(A) \in Z(R)$, then A is both a right and left zero divisor in $M_{n\times n}(R)$.

Conversely, suppose A is a left zero divisor in $M_{n\times n}(R)$. Then $AB = O$ for some nonzero $B \in M_{n\times n}(R)$. Suppose $B = (\xi_1 \mid \cdots \mid \xi_n)$ is a column partition of B. Since B is not zero, some ξ_i is a nonzero vector in R^n. $O = AB = (A\xi_1 \mid \cdots \mid A\xi_n)$ implies $A\xi_i = O$ for all $i = 1, \ldots, n$. In particular, the equation $AX = O$ has a nontrivial solution. Theorem 5.3 implies $\text{rk}(A) < n$, and then $\det(A) \in Z(R)$ by 4.11e.

If A is a right zero divisor, then A^t is a left zero divisor. Thus, $\det(A) = \det(A^t) \in Z(R)$. ∎

Theorem 9.1 implies the set of left zero divisors of $M_{n\times n}(R)$ is precisely the same as the set of right zero divisors in $M_{n\times n}(R)$. Either of these sets is just $Z(M_{n\times n}(R))$. If R is a field, or for that matter any integral domain, then $Z(R) = (0)$. Hence, we have the following corollary to Theorem 9.1.

Corollary 9.2 Let R be an integral domain and $A \in M_{n\times n}(R)$. Then $A \in Z(M_{n\times n}(R))$ if and only if $\det(A) = 0$. ∎

It follows from Corollary 7.25 that the inverse of an invertible matrix A is a polynomial in A. There is a similar result for zero divisors. If A is a zero divisor in $M_{n\times n}(R)$, then there exists a polynomial $g(X) \in R[X]$ such that $g(A) \neq O$, and $Ag(A) = g(A)A = O$. Thus, A is a zero divisor in $M_{n\times n}(R)$ if and only if A is a zero divisor in the commutative subalgebra $R[A]$. In order to prove these statements, we need the following preliminary result.

Lemma 9.3 Let $C, D \in M_{n\times n}(R)$, and let $x \in R$. If $xC = xD$, then $x \det(C) = x \det(D)$.

Proof. Partition both C and D into rows: $C = (\lambda_1; \ldots; \lambda_n)$ and $D = (\mu_1; \ldots; \mu_n)$. Then $xC = xD$ implies $x\lambda_i = x\mu_i$ for all $i = 1, \ldots, n$. Using the fact that the determinant is a multilinear function of its rows, we have

$$\begin{aligned} x \det(C) &= x \det(\lambda_1; \ldots; \lambda_n) = \det(x\lambda_1; \lambda_2; \ldots; \lambda_n) \\ &= \det(x\mu_1; \lambda_2; \ldots; \lambda_n) = x \det(\mu_1; \lambda_2; \ldots; \lambda_n) \\ &= \det(\mu_1; x\lambda_2; \ldots; \lambda_n) = \det(\mu_1; x\mu_2; \ldots; \lambda_n) \\ &= x \det(\mu_1; \mu_2; \lambda_3; \ldots; \lambda_n) = \cdots \\ &= x \det(\mu_1; \ldots; \mu_n) = x \det(D). \end{aligned}$$ ∎

We can now state the following theorem.

Theorem 9.4 $A \in Z(M_{n\times n}(R))$ if and only if $A \in Z(R[A])$.

Proof. Since $R[A] \subseteq M_{n\times n}(R)$, the implication from right to left in Theorem 9.4 is obvious. The theorem is also clear if $n = 1$. Hence, we assume $n \geq 2$, and $A \in Z(M_{n\times n}(R))$.

We first give the proof in the special case $n = 2$. Let $C_A(X) = X^2 + a_1X + a_2$. Then $a_2 = \det(A) \in Z(R)$ by Theorem 9.1. Hence, there exists an element $b \in R^*$ such that $ba_2 = 0$. By the Cayley-Hamilton theorem, $O = bC_A(A) = bA^2 + ba_1A = A(bA + ba_1I_2)$. If $bA + ba_1I_2 \neq O$, then $A \in Z(R[A])$.

Suppose $bA + ba_1I_2 = O$. Then $bA = b(-a_1I_2)$. Lemma 9.3 implies $b\det(A) = b\det(-a_1I_2) = ba_1^2$. But $b\det(A) = 0$. Therefore, $ba_1^2 = 0$. Set $i = \min\{j \mid ba_1^j = 0\}$. Then $i = 1$ or 2. Set $b' = ba_1^{i-1}$. Thus, $b' = b$ if $i = 1$, and $b' = ba_1$ if $i = 2$. In either case, $b' \neq 0$ and

$$O = a_1^{i-1}(bA^2 + ba_1A) = b'A^2 + ba_1^iA = b'A^2 = (b'A)A.$$

If $b'A = O$, then $A \in Z(R[A])$. If $b'A \neq O$, then $(b'A)A = O$ implies $A \in Z(R[A])$. Thus, in either case, A is a zero divisor in $R[A]$.

We can now assume $n \geq 3$. The proof proceeds in the same manner as before. Let $C_A(X) = X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n$. Then $\det(A) = \pm a_n \in Z(R)$ by Theorem 9.1. Therefore, there exists an element $b \in R^*$ such that $ba_n = 0$. Again by the Cayley-Hamilton theorem, we have $O = bC_A(A) = bA(A^{n-1} + a_1A^{n-2} + \cdots + a_{n-1})$. If $b(A^{n-1} + a_1A^{n-2} + \cdots + a_{n-1}) \neq O$, then $A \in Z(R[A])$. Hence, we can assume $b(A^{n-1} + a_1A^{n-2} + \cdots + a_{n-1}) = O$. We then have the following equation in $M_{n\times n}(R)$:

$$\textbf{9.5}\quad bA(A^{n-2} + a_1A^{n-3} + \cdots + a_{n-2}) = -ba_{n-1}I_n$$

Lemma 9.3 implies

$$b\det(A)\det(A^{n-2} + a_1A^{n-3} + \cdots + a_{n-2}) = b(-1)^n a_{n-1}^n$$

Since $b\det(A) = 0$, $ba_{n-1}^n = 0$.

Set $i = \min\{j \mid ba_{n-1}^j = 0\}$. Then $1 \leq i \leq n$. Set $b' = ba_{n-1}^{i-1}$. Then $b' = \neq 0$, and $b'a_n = b'a_{n-1} = 0$. Multiplying equation 9.5 with a_{n-1}^{i-1}, we have

$$\textbf{9.6}\quad A(b'(A^{n-2} + a_1A^{n-3} + \cdots + a_{n-2})) = O$$

If $b'(A^{n-2} + a_1A^{n-3} + \cdots + a_{n-2}) \neq O$, then equation 9.6 implies $A \in Z(R[A])$. If $b'(A^{n-2} + a_1A^{n-3} + \cdots + a_{n-2}) = O$, then we have

$$\textbf{9.7}\quad b'A(A^{n-3} + a_1A^{n-4} + \cdots + a_{n-3}) = -b'a_{n-2}I_n$$

This is an equation like 9.5 only one degree lower. We can now repeat this argument. After a finite number of repetitions, we get $A(f(A)) = O$ for some $f(A) \in R[A]^*$, or $cA = O$ for some $c \in R^*$. In either case, $A \in Z(R[A])$. ∎

We can extend the result in Theorem 9.4 slightly as follows:

Corollary 9.8 Let $A \in M_{n\times n}(R)$ and $g(X) \in R[X]$. Then $g(A) \in Z(M_{n\times n}(R))$ if and only if $g(A) \in Z(R[A])$.

Proof. Set $B = g(A)$. If $B \in Z(M_{n\times n}(R))$, then $B \in Z(R[B])$ by Theorem 9.4. Since $R[B] \subseteq R[A]$, $B = g(A) \in Z(R[A])$. The implication in the other direction is trivial. ∎

We now return to the null ideal N_A of A. The reader will recall that N_A is the kernel of the R-algebra homomorphism $\vartheta_A : R[X] \mapsto R[A]$ given by $\vartheta_A(f(X)) = f(A)$. We have seen in Corollary 7.38 that the minimal primes of N_A are precisely the same as the minimal primes of the principal ideal $(C_A(X))$ in $R[X]$. We can now argue the maximal primes belonging to N_A are precisely the same as the maximal primes belonging to $(C_A(X))$. We need the following result.

Theorem 9.9 Let $A \in M_{n\times n}(R)$, and let $g(X) \in R[X]$. Then $g \in Z(R[X]/N_A)$ if and only if $g \in Z(R[X]/(C_A(X)))$.

Proof. The result is obvious if $g(X) = 0$. Hence, we may assume $g(X)$ is a nonzero polynomial in $R[X]$. It follows from 7.29 that ϑ_A induces an isomorphism $\overline{\vartheta}_A : R[X]/N_A \cong R[A]$. Under this isomorphism, $g \in Z(R[X]/N_A)$ if and only if $g(A) \in Z(R[A])$. The matrix $g(A)$ is a zero divisor (possibly zero) in $R[A]$ if and only if $\det(g(A)) \in Z(R)$. This follows from Theorem 9.1 and Corollary 9.8. By Theorem 8.54, $\mathfrak{R}(C_A(X),g(X)) = \det(g(A))$. Thus, $g(A) \in Z(R[A])$ if and only if $\mathfrak{R}(C_A(X),g(X)) \in Z(R)$. But Theorem 8.14 implies $\mathfrak{R}(C_A(X),g(X)) \in Z(R)$ if and only if $g(X) \in Z(R[X]/(C_A(X)))$. Putting all this together, we have $g(X)$ is a zero divisor in $R[X]/N_A$ if and only if $g(X)$ is a zero divisor in $R[X]/(C_A(X))$. ∎

Theorem 9.9 says $Z(R[X]/N_A) = Z(R[X]/(C_A))$. In particular, suppose $\mathfrak{P}$ is a maximal prime belonging to N_A. Then $\mathfrak{P} \subseteq Z(R[X]/N_A)$, and $\mathfrak{P}$ is maximal among the ideals in $Z(R[X]/N_A)$. Since $Z(R[X]/N_A) = Z(R[X]/(C_A))$, $\mathfrak{P} \subseteq Z(R[X]/(C_A))$ and $\mathfrak{P}$ is maximal among the ideals in $Z(R[X]/(C_A))$. Thus, $\mathfrak{P}$ is a maximal prime belonging to (C_A). The converse statement is also clear. Hence we have the following corollary to Theorem 9.9.

Corollary 9.10 Let $A \in M_{n\times n}(R)$. The maximal primes belonging to N_A are precisely the same as the maximal primes belonging to (C_A). ∎

Combining the results from Corollary 7.38 and Corollary 9.10, we see N_A and $(C_A(X))$ are fairly close as ideals in $R[X]$ in the sense that the minimal primes and the maximal primes belonging to each ideal are the same.

EXERCISES

1. Let F be a field. Show every matrix in $M_{n\times n}(F)$ is either a zero divisor or invertible. Is this true for an arbitrary commutative ring R?
2. Let p be a positive prime in $\mathbb{Z}$, and set $T = M_{n\times n}(\mathbb{Z}/p\mathbb{Z})$. Compute the cardinality of $Z(T)$. (See Exericse 15 in Chapter 2.)
3. Find an example of a noncommutative ring T and an element $x \in T$ such that x is a left zero divisor but not a right zero divisor in T.
4. Let $R = \mathbb{Z}/6\mathbb{Z}$. For each matrix A listed below, show A is a zero divisor and find a polynomial $g(X) \in R[X]$ such that $g(A) \neq O$ and $Ag(A) = O$:

(a) $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ (b) $A = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix}$ (c) $A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 3 & 4 \\ 1 & 2 & 5 \end{bmatrix}$

5. Compute the minimal primes of N_A and the maximal primes belonging to N_A for each matrix A given in Exercise 4.
6. Let R be a commutative ring. If $\mathfrak{A}$ and $\mathfrak{B}$ are ideals in R, then the quotient of $\mathfrak{A}$ and $\mathfrak{B}$ is the set $\mathfrak{A} : \mathfrak{B} = \{x \in R \mid x\mathfrak{B} \subseteq \mathfrak{A}\}$. Prove the following facts about quotients:
   (a) $\mathfrak{A} : \mathfrak{B}$ is an ideal in R.
   (b) $\mathfrak{A} : \mathfrak{B} = R$ if and only if $\mathfrak{B} \subseteq \mathfrak{A}$.
   (c) $(\cap_{i=1}^n \mathfrak{A}_i) : \mathfrak{B} = \cap_{i=1}^n (\mathfrak{A}_i : \mathfrak{B})$.
   (d) $\mathfrak{A} : (\Sigma_{i=1}^n \mathfrak{B}_i) = \cap_{i=1}^n (\mathfrak{A} : \mathfrak{B}_i)$.
   (e) $\mathfrak{A} : \mathfrak{B}\mathfrak{C} = (\mathfrak{A} : \mathfrak{B}) : \mathfrak{C}$.
7. Let $A \in M_{n\times n}(R)$. Prove the following theorem of N.McCoy:

$$N_A = (C_A(X)) : I_{n-1}(XI_n - A)$$

8. Let $A \in M_{n\times n}(R)$ and $g(X) \in R[X]$. Prove the following theorem of N. McCoy: $g(A)$ is nilpotent if and only if $g(X) \in \sqrt{N_A}$.
9. Let $R = \mathbb{Z}/12\mathbb{Z}$. Use Theorem 8.14 to show $g(X) = 2X + 5$ is not a zero divisor in $R[X]/(f)$ where $f(X) = X^2 + 3$.
10. Use Exercise 9 to show $g(X) = 2X + 5$ is contained in no ideal belonging to N_A where

$$A = \begin{bmatrix} 3 & 6 \\ 0 & 9 \end{bmatrix} \in M_{2\times 2}(\mathbb{Z}/12\mathbb{Z}).$$

10

Finitely Generated Modules and Local Rings

Here and in the next two chapters, we will discuss the basic theorems in commutative ring theory which will be used throughout the rest of the text. In this chapter, we concentrate on modules which satisfy the ascending chain condition and their connections with finitely generated modules. At the end of this chapter, we will say a few words about local rings. As usual, R denotes a commutative ring. Let M be an R-module.

Definition 10.1 M is said to satisfy the ascending chain condition if every strictly ascending chain of submodules of M, $N_1 < N_2 < N_3 < \cdots$, is finite.

Clearly, M satisfies the ascending chain condition if for every infinite ascending chain of submodules $N_1 \subseteq N_2 \subseteq N_3 \subseteq \cdots$ of M there exists a positive integer r such that $N_i = N_r$ for all $i \geq r$. The ascending chain condition is equivalent to two other important conditions on the submodules of M.

Lemma 10.2 Let M be an R-module. Then the following statements are equivalent:

(a) M satisfies the ascending chain condition.
(b) Every nonempty set of submodules of M has a maximal element with respect to inclusion.
(c) Every submodule of M is a finitely generated R-module.

Proof. Suppose M satisfies the ascending chain condition. Let $\mathcal{F}$ be a nonempty set of submodules of M. Suppose $\mathcal{F}$ has no maximal element with respect to inclusion. Then if $N_1 < N_2 < \cdots < N_r$ is any strictly ascending chain of submodules from $\mathcal{F}$, there exists an $N_{r+1} \in \mathcal{F}$ such that $N_1 < N_2 < \cdots < N_r < N_{r+1}$. To see this, we merely note that N_r is not a maximal element of $\mathcal{F}$, and therefore there exists an $N_{r+1} \in \mathcal{F}$ such that $N_r < N_{r+1}$. By ordering chains of submodules of $\mathcal{F}$ in the obvious way and using Zorn's lemma, we can construct an infinite, strictly ascending chain $N_1 < N_2 < \cdots$ of submodules from $\mathcal{F}$. (See Exercise 1 for more details.) This is impossible since M satisfies the ascending chain condition. We conclude that every nonempty set of submodules of M contains a maximal element. Thus, $(a) \Rightarrow (b)$.

Suppose M satisfies (b). Let $N_1 \subseteq N_2 \subseteq N_3 \subseteq \cdots$ be an ascending chain of submodules of M. Set $\mathcal{F} = \{N_i \mid i = 1, 2, 3, \ldots\}$. By (b), $\mathcal{F}$ contains a maximal element. Suppose N_r is a maximal element of $\mathcal{F}$. Then clearly $N_i = N_r$ for all $i \geq r$. Thus, M satisfies the ascending chain condition. Therefore, $(b) \Rightarrow (a)$.

Suppose every submodule of M is finitely generated. Let $N_1 \subseteq N_2 \subseteq N_3 \subseteq \cdots$ be an ascending chain of submodules of M. Then $\bigcup_{i=1}^{\infty} N_i$ is a submodule of M. By hypothesis, $\bigcup_{i=1}^{\infty} N_i$ is a finitely generated R-module. Suppose $\{m_1, \ldots, m_n\}$ is a set of generators of $\bigcup_{i=1}^{\infty} N_i$. Since the N_i ascend, there exists an $r > 0$ such that $m_1, \ldots, m_n \in N_r$. But then $\bigcup_{i=1}^{\infty} N_i = N_r$, and, in particular, $N_i = N_r$ for all $i \geq r$. Thus, (c) $\Rightarrow$ (a).

Finally, suppose M satisfies (b). Let N be a submodule of M. Let $\mathcal{F}$ be the set of all submodules of N which are finitely generated. $\mathcal{F} \neq \emptyset$, since $(0) \in \mathcal{F}$. By (b), $\mathcal{F}$ contains a maximal element N'. Let $m \in N$. Then $N' + Rm \subseteq N$, and $N' + Rm$ is finitely generated since N' is. Therefore, $N' + Rm \in \mathcal{F}$. The maximality of N' in $\mathcal{F}$ implies $N' = N' + Rm$. In particular, $m \in N'$. Thus, $N = N'$, and N is a finitely generated submodule of M. We have now shown (b) $\Rightarrow$ (c). This completes the proof of the lemma. ∎

One important application of Lemma 10.2 occurs when $M = R$, the ring itself. If $M = R$, then the R-submodules of R are just the ideals of R. Thus, the R-module R satisfies the ascending chain condition if and only if every ideal in R is finitely generated. In other words, the R-module R satisfies the ascending chain condition if and only if R is a Noetherian ring. We have already noted many examples of Noetherian rings. Any PIR is Noetherian. Thus, all the rings listed in 1.7 together with their homomorphic images are Noetherian rings. In particular, these rings satisfy the ascending chain condition on ideals.

Another important class of Noetherian rings is provided by the following famous theorem of D. Hilbert.

Theorem 10.3 (Hilbert Basis Theorem) Let R be a Noetherian ring. For any indeterminate X, the polynomial ring $R[X]$ is also Noetherian.

Proof. Let $\mathfrak{A}$ be an ideal in $R[X]$. We must show $\mathfrak{A}$ is a finitely generated ideal. We can assume $\mathfrak{A} \neq (0)$. Suppose $\mathfrak{A}$ is not finitely generated. We can then construct an infinite sequence of nonzero polynomials $f_1(X), f_2(X), \ldots$ in $\mathfrak{A}$ in the following way: $\mathfrak{A} \neq (0)$. Hence there exists a nonzero $f_1(X) \in \mathfrak{A}$ such that the degree of f_1, $\partial(f_1)$, is as small as possible. So, $\partial(f_1) = \min\{\partial(g) \mid g \in \mathfrak{A} - (0)\}$. Suppose we have constructed $f_1, \ldots, f_k$. Since $\mathfrak{A}$ is not finitely generated, $\mathfrak{A} > (f_1, \ldots, f_k)$, the ideal generated by $f_1, \ldots, f_k$. Let $f_{k+1}(X)$ be a polynomial in $\mathfrak{A} - (f_1, \ldots, f_k)$ of minimal degree. In this way, we construct a sequence of nonzero polynomials $f_1, f_2, \ldots$ in $\mathfrak{A}$ such that

$$\partial(f_{i+1}) = \min\{\partial(g) \mid g \in \mathfrak{A} - (f_1, \ldots, f_i)\}$$

for all $i \geq 1$

Let $n(i) = \partial(f_i)$, and let r_i denote the leading coefficient of $f_i(X)$ for each $i \geq 1$. Then $f_i(X) = r_i X^{n(i)} +$ lower-degree terms. Since $f_{i+1} \in \mathfrak{A} - (f_1, \ldots, f_i)$, $f_{i+1} \in \mathfrak{A} - (f_1, \ldots, f_{i-1})$. In particular

$$n(i) = \min\{\partial(g) \mid g \in \mathfrak{A} - (f_1, \ldots, f_{i-1})\} \leq n(i+1)$$

Thus, $n(i) \leq n(i+1)$ for all $i \geq 1$. We claim

10.4 $(r_1) < (r_1, r_2) < (r_1, r_2, r_3) < \cdots$

is a strictly ascending chain of ideals in R.

To see this, suppose $(r_1, \ldots, r_k) = (r_1, \ldots, r_{k+1})$ for some $k \geq 1$. Then $r_{k+1} = \sum_{i=1}^{k} a_i r_i$ for some $a_1, \ldots, a_k \in R$. Consider the polynomial

$$g(X) = f_{k+1}(X) - \sum_{i=1}^{k} a_i X^{n(k+1)-n(i)} f_i(X).$$

Since $f_{k+1} \in \mathfrak{A} - (f_1, \ldots, f_k)$, $g \in \mathfrak{A} - (f_1, \ldots, f_k)$. Since $\sum_{i=1}^{k} a_i r_i = r_{k+1}$, $\partial(g) < \partial(f_{k+1})$. The definition of f_{k+1} implies this last inequality is impossible. Thus, $(r_1, \ldots, r_k) < (r_1, \ldots, r_{k+1})$, and the chain of ideals in 10.4 is strictly increasing.

Since R is a Noetherian ring, a chain like that in 10.4 is impossible. We conclude that $\mathfrak{A}$ is finitely generated. Thus, $R[X]$ is a Noetherian ring. ∎

Let R be a Noetherian ring and $X_1, \ldots, X_n$ indeterminates over R. Theorem 10.3 together with a simple induction argument imply $R[X_1, \ldots, X_n]$ is a Noetherian ring. Since the ideals in any homomorphic image $R/\mathfrak{A}$ of R are in a one-to-one, order-preserving correspondence with the ideals of R which contain $\mathfrak{A}$, any homomorphic image of a Noetherian ring is Noetherian. In particular, any ring of the form $R[X_1, \ldots, X_n]/\mathfrak{A}$ is Noetherian when R is Noetherian.

Rings of the form $R[X_1, \ldots, X_n]/\mathfrak{A}$ are called finitely generated (commutative) R-algebras. Hence, we have the following corollary to Theorem 10.3.

Corollary 10.5 Let R be a Noetherian ring. Then any (commutative) finitely generated R-algebra is also a Noetherian ring. ∎

For example, $\mathbb{Z}[X_1, \ldots, X_n]/\mathfrak{A}$ and $F[X_1, \ldots, X_n]/\mathfrak{A}$ are Noetherian rings.

An R-module M is certainly a submodule of itself. Thus, Lemma 10.2 implies that if M satisfies the ascending chain condition, then M is a finitely generated R-module. The converse of this statement is not true in general. A finitely generated R-module need not satisfy the ascending chain condition. Any non-Noetherian ring will furnish an example.

Example 10.6 Let $X_1, X_2, \ldots$ be a countable number of indeterminates over the field F. Set $R = F[X_1, X_2, \ldots]$, the polynomial ring over F in the variables $X_1, X_2, \ldots$. Let $M = R$. Then $\{1\}$ is an R-module basis of M. Thus, M is a finitely generated R-module. Set $\mathfrak{A} = (X_1, X_2, \ldots)$. Clearly, $\mathfrak{A}$ is not finitely generated. Thus, by Lemma 10.2, M fails to satisfy the ascending chain condition. ∎

Submodules and homomorphic images of a module satisfying the ascending chain condition also satisfy the ascending chain condition. The converse of this statement is also true.

Theorem 10.7 Let M be an R-module and N an R-submodule of M. M satisfies the ascending chain condition if and only if both N and M/N satisfy the ascending chain condition.

Proof. We need only show if N and M/N satisfy the ascending chain condition, then so does M. Let P be a submodule of M. Then $P \cap N$ and $(P + N)/N$ are submodules of N and M/N, respectively. By Lemma 10.2, both of these submodules are finitely generated. Let $\{m_1, \ldots, m_n\}$ be a basis of $P \cap N$, and let $\{\bar{y}_1, \ldots, \bar{y}_k\}$ be a basis of $(P + N)/N$. Since P maps onto $(P + N)/N$ via the natural homomorphism, there exist elements $y_1, \ldots, y_k \in P$ such that $\bar{y}_i = y_i + N$ for all $i = 1, \ldots, k$. We claim that $\Delta = \{m_1, \ldots, m_n, y_1, \ldots, y_k\}$ is a basis of P.

Let $p \in P$. Then $\bar{p} = p + N = \sum_{i=1}^{k} r_i \bar{y}_i$ in $(P + N)/N$. Here $r_1, \ldots, r_k$ are elements in R. Back in P, we have

$$p - \sum_{i=1}^{k} r_i y_i \in P \cap N = Rm_1 + \cdots + Rm_n$$

Therefore, there exist ring elements $s_1, \ldots, s_n \in R$ such that $p - \sum_{i=1}^{k} r_i y_i = \sum_{j=1}^{n} s_j m_j$. In particular, $p = \sum_{i=1}^{k} r_i y_i + \sum_{j=1}^{n} s_j m_j$. Since p is arbitrary, we conclude Δ is a basis of P. Consequently, P is finitely generated.

Since P is an arbitrary submodule of M, every submodule of M is finitely generated. It follows from Lemma 10.2 that M satisfies the ascending chain condition. ∎

There are several important corollaries to Theorem 10.7.

Corollary 10.8 Suppose R is a Noetherian ring. Then for any positive integer n, the R-module R^n satisfies the ascending chain condition.

Proof. Let $\varepsilon = \{\varepsilon_1, \ldots, \varepsilon_n\}$ denote the canonical basis of R^n. Then $R\varepsilon_1 \cong R$, and $R^n/R\varepsilon_1 \cong R^{n-1}$ as R-modules. Since R is Noetherian, $R\varepsilon_1$ satisfies the ascending chain condition. By induction on n, we can assume R^{n-1} satisfies the ascending chain condition. In particular, $R^n/R\varepsilon_1$ satisfies the ascending chain condition. Theorem 10.7 now implies R^n satisfies the ascending chain condition. ∎

Before stating our next corollary, we need the following definitions from homological algebra.

Definition 10.9 (a)

$$(*) : (0) \mapsto N \overset{g}{\mapsto} P \overset{f}{\mapsto} M \mapsto (0)$$

is a five-term complex of R-modules if N, P, and M are R-modules, g and f are R-module homomorphisms (i.e., $g \in \mathrm{Hom}_R(N, P)$ and $f \in \mathrm{Hom}_R(P, M)$), and $fg = 0$.

(b) The complex $(*)$ in (a) is said to be right exact if f is surjective and $\mathrm{Im}(g) = \mathrm{Ker}(f)$.

(c) The complex $(*)$ in (a) is called a short exact sequence (of R-modules) if $(*)$ is right exact and g is injective.

In Definition 10.9, (0) denotes the zero module. Thus, $(*)$ is a sequence of five R-modules, (0), N, P, M, and (0). The maps $(0) \mapsto N$ and $M \mapsto (0)$ in $(*)$ are the obvious R-module homomorphisms sending $0 \mapsto 0$ (in $(0) \mapsto N$) and $m \mapsto 0$ for all $m \in M$ (in $M \mapsto (0)$). This explains why $(*)$ is called a five-term complex. The word complex means successive composites of maps are zero. Since the end terms in $(*)$ are (0), this simply means $fg = 0$. The definitions imply a five-term complex of R-modules of the form

$$(*) : (0) \mapsto N \overset{g}{\mapsto} P \overset{f}{\mapsto} M \mapsto (0)$$

is a short exact sequence if and only if the following three conditions are satisfied:

(a) g is injective (i.e., $\mathrm{Ker}(g) = (0)$).

(b) $\mathrm{Im}(g) = \mathrm{Ker}(f)$.

(c) f is surjective (i.e., $\mathrm{Im}(f) = M$).

One way to manufacture short exact sequences is as follows: Suppose M is an R-module and N is a submodule of M. Let $\nu : M \mapsto M/N$ denote the natural map given by $\nu(m) = m + N \in M/N$. Let $\iota : N \mapsto M$ be the inclusion map of N into M. Then clearly,

$$(0) \mapsto N \overset{\iota}{\mapsto} M \overset{\nu}{\mapsto} M/N \mapsto (0)$$

is a short exact sequence.

Now suppose M is a finitely generated R-module. Let $\Gamma = \{m_1, \ldots, m_n\}$ be an R-module basis of M. (If $M = (0)$, we take $n = 1$ and $m_1 = 0$.) Let $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_n\}$ be any free R-module basis of R^n. The two bases Γ and $\boldsymbol{\lambda}$ determine a unique R-module homomorphism $f : R^n \mapsto M$ which sends λ_i to m_i for all $i = 1, \ldots, n$. This map is defined as follows: $f(\sum_{i=1}^n r_i\lambda_i) = \sum_{i=1}^n r_i m_i$. Since $\boldsymbol{\lambda}$ is a free R-module basis of R^n, the reader can easily check that f is well defined and $f(\lambda_i) = m_i$ for all $i = 1, \ldots, n$. Since Γ is a basis of M, the R-module homomorphism $f : R^n \mapsto M$ is surjective.

Let $K = \mathrm{Ker}(f)$. Then K is an R-submodule of R^n, and

$$\textbf{10.10} \quad (0) \mapsto K \overset{\iota}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$$

is a short exact sequence of R-modules. The map $\iota : K \mapsto R^n$ in 10.10 is the inclusion map of K into R^n. Henceforth, we will drop ι from the notation.

If R is a Noetherian ring, then we can say a bit more about K. By Corollary 10.8, R^n satisfies the ascending chain condition. Thus, Lemma 10.2 implies K is a finitely generated R-module. Suppose $\Delta = \{\delta_1, \ldots, \delta_m\}$ is an R-module basis of K. Let $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_m\}$ be a free R-module basis of R^m (If $K = (0)$, we take $m = 1$ and $\delta_1 = 0$.) The two R-module bases Δ and $\boldsymbol{\mu}$ define a unique R-module homomorphism $g : R^m \mapsto K$ sending μ_i to δ_i for all $i = 1, \ldots, m$. Obviously, g is defined by the following equation: $g(\sum_{i=1}^m r_i\mu_i) = \sum_{i=1}^m r_i\delta_i$. The homomorphism g is clearly surjective, and thus $\mathrm{Im}(g) = K = \mathrm{Ker}(f)$. In particular,

$$\textbf{10.11} \quad R^m \overset{g}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$$

is a right exact sequence of R-modules. Notice that we have dropped the symbols "$(0) \mapsto$" from the left side of the sequence in 10.11.

Any right exact sequence of the form

$$F_1 \mapsto F_0 \mapsto M \mapsto (0)$$

in which both F_1 and F_0 are free R-modules is called a presentation of M. Thus, the sequence in 10.11 is a presentation of the R-module M. The module M is said to be finitely presented if M admits a presentation $F_1 \mapsto F_0 \mapsto M \mapsto (0)$ in which both F_1 and F_0 are finitely generated, free R-modules. If F_1 and F_0 are both finitely generated, free R-modules, then the right exact sequence $F_1 \mapsto F_0 \mapsto M \mapsto (0)$ is called a finite presentation of M. Thus, the sequence in 10.11 is a finite presentation of M. We have now proved our second corollary to Theorem 10.7.

Corollary 10.12 Any finitely generated module over a Noetherian ring is finitely presented. ∎

We note in passing that a finitely presented R-module M may admit no right exact sequence like that in 10.11 in which g is also injective. In other words, there may be no short exact sequences of the form

$$(0) \mapsto R^m \mapsto R^n \mapsto M \mapsto (0)$$

The question as to whether such short exact sequences exist or not depends on the projective dimension of the module M. We will discuss this topic briefly in Chapter 13.

Let us return to the finite presentation of M given in 10.11. There is certainly nothing unique about this sequence. As we vary our choices of Γ, $\boldsymbol{\lambda}$, Δ, and $\boldsymbol{\mu}$, we get various finite presentations of M. Of course, if R is not Noetherian, a finitely generated R-module M may admit no finite presentation. To see this, return to Example 10.6. Suppose we set $M = R/\mathfrak{A}$ in Example 10.6. Then M is a finitely generated R-module with basis $\Gamma = \{1 + \mathfrak{A}\}$. Since $\mathfrak{A} = (X_1, X_2, \ldots)$ is not finitely generated, the reader can easily argue that M is not finitely presented.

Suppose M has a finite presentation as in 10.11. We use the same notation Γ, $\boldsymbol{\lambda}$, Δ, $\boldsymbol{\mu}$, f, and g as used before. There is a natural $m \times n$ matrix $C \in M_{m \times n}(R)$ which can be associated with the data Γ, $\boldsymbol{\lambda}$, Δ, and $\boldsymbol{\mu}$. Each vector $g(\mu_i) = \delta_i$ in K is a unique linear combination of the vectors in $\boldsymbol{\lambda}$. Suppose we write

$$\textbf{10.13} \quad g(\mu_i) = \delta_i = \sum_{j=1}^{n} c_{ij}\lambda_j \quad \text{for } i = 1, \ldots, m$$

The c_{ij} in equation 10.13 are elements from R. Set $C = (c_{ij}) \in M_{m \times n}(R)$. The $m \times n$ matrix C is called a relations matrix of M.

The reason for this name is that the rows of C generate all relations among the generators $m_1, \ldots, m_n$ of M. By a relation among the generators $m_1, \ldots, m_n$, we mean an n-tuple $(r_1, \ldots, r_n) \in M_{1 \times n}(R)$ such that $r_1m_1 + \cdots + r_nm_n = 0$. Since Γ is not necessarily a free R-module basis of M (M may not even be free), there could be nontrivial relations $(r_1, \ldots, r_n) \in M_{1 \times n}(R)^*$ among

the generators $m_1, \ldots, m_n$. We claim $\mathrm{RS}(C)$ is the complete set of relations among the generators $m_1, \ldots, m_n$.

Consider the ith row $(c_{i1}, \ldots, c_{in})$ of C. By equation 10.13,

$$g(\mu_i) = \delta_i = \sum_{j=1}^{n} c_{ij}\lambda_j \in K = \mathrm{Ker}(f)$$

Since f is an R-module homomorphism

$$0 = f(\delta_i) = \sum_{j=1}^{n} c_{ij} f(\lambda_j) = \sum_{j=1}^{n} c_{ij} m_j.$$

Thus, $(c_{i1}, \ldots, c_{in})$ is a relation among $m_1, \ldots, m_n$. If some entry in the row $(c_{i1}, \ldots, c_{in})$ is nonzero, then the ith row of C is a nontrivial relation among the generators $m_1, \ldots, m_n$. At any rate, each row of C determines a relation among the generators $m_1, \ldots, m_n$.

Conversely, suppose $(r_1, \ldots, r_n)$ is a relation among $m_1, \ldots, m_n$. Thus, $r_1 m_1 + \cdots + r_n m_n = 0$. Since $\sum_{j=1}^{n} r_j m_j = 0$,

$$\sum_{j=1}^{n} r_j \lambda_j \in \mathrm{Ker}(f) = \mathrm{Im}(g) = Rg(\mu_1) + \cdots + Rg(\mu_m).$$

Therefore, $\sum_{j=1}^{n} r_j \lambda_j = \sum_{i=1}^{m} s_i g(\mu_i)$ for some $s_1, \ldots, s_m \in R$. Using equation 10.13, we have

$$\sum_{j=1}^{n} r_j \lambda_j = \sum_{i=1}^{m} s_i g(\mu_i) = \sum_{i=1}^{m} s_i \left(\sum_{j=1}^{n} c_{ij} \lambda_j \right)$$

$$= \sum_{j=1}^{n} \left(\sum_{i=1}^{m} s_i c_{ij} \right) \lambda_j$$

Since $\{\lambda_1, \ldots, \lambda_n\}$ is a free R-module basis of R^n, we have $r_j = \sum_{i=1}^{m} s_i c_{ij}$ for each $j = 1, \ldots, n$. This last equation implies $(r_1, \ldots, r_n) = s_1 \mathrm{Row}_1(C) + \cdots + s_m \mathrm{Row}_m(C)$. Thus, any relation $(r_1, \ldots, r_n)$ among $m_1, \ldots, m_n$ is a vector in $\mathrm{RS}(C)$.

Notice that the relations matrix C constructed from 10.11 depends on our choices of Γ, $\boldsymbol{\lambda}$, Δ, and $\boldsymbol{\mu}$. Thus, there are many relations matrices for the same finitely presented R-module M. If we choose $\boldsymbol{\lambda} = \boldsymbol{\varepsilon}$ and $\boldsymbol{\mu} = \boldsymbol{\varepsilon}$, the canonical bases in R^n and R^m, respectively, then there is a simple relationship between the homomorphism g and the matrix C. We have

10.14 $g(\xi) = C^t \xi$ for all $\xi \in R^m$.

Equation 10.14 follows easily from 10.13. We leave this as an exercise at the end of this chapter.

We will put these ideas to work in Chapter 13 and beyond. For now, we summarize this discussion with the following remark: Any finitely generated module M over a Noetherian ring R has a relations matrix $C \in M_{m \times n}(R)$ (for some choice of m and n). The rows of C determine the complete set of relations among a set of generators of M.

We finish this Chapter with Nakayama's lemma and some applications to local rings.

Lemma 10.15 (Nakayama's Lemma) Let R be a commutative ring, and let M be a finitely generated R-module. Suppose $M = \mathfrak{A}M$ for some ideal $\mathfrak{A} \subseteq J(R)$. Then $M = (0)$.

Proof. The reader will recall that $J(R)$ is the Jacobson radical of of R. Since $\mathfrak{A} \subseteq J(R)$, every element of $\mathfrak{A}$ is quasi-regular. Suppose $M \neq (0)$. Since M is a finitely generated R-module, there exists a positive integer n such that M has a basis consisting of n elements but M has no basis with fewer than n elements. Suppose $\{m_1, \ldots, m_n\}$ is one basis of M containing n elements. Since $M = \mathfrak{A}M$, $m_n = a_1m_1 + a_2m_2 + \cdots + a_nm_n$ for some choice of $a_1, \ldots, a_n \in \mathfrak{A}$. Thus, $(1 - a_n)m_n = a_1m_1 + \cdots + a_{n-1}m_{n-1}$.

Since $\mathfrak{A}$ is a quasi-regular ideal in R, $1 - a_n \in U(R)$. Therefore,

$$m_n = [(1 - a_n)^{-1}a_1]m_1 + \cdots + [(1 - a_n)^{-1}a_{n-1}]m_{n-1}.$$

But then $\{m_1, \ldots, m_{n-1}\}$ is a basis of M. This is impossible since n is the smallest number of elements which can generate M. We conclude that $M = (0)$. ∎

There are a couple of corollaries of Nakayama's lemma which are very useful.

Corollary 10.16 Suppose M is a finitely generated R-module. Let $\mathfrak{A}$ be an ideal of R with $\mathfrak{A} \subseteq J(R)$. Let N be a submodule of M. If $M = N + \mathfrak{A}M$, then $M = N$.

Proof. Since M is a finitely generated R-module, so is M/N. $\mathfrak{A}(M/N) = (N + \mathfrak{A}M)/N = M/N$ by hypothesis. Nakayama's lemma then implies $M/N = (0)$, that is, $M = N$. ∎

Corollary 10.16 is particularly useful when R is a local ring. The reader will recall that a commutative ring R is called a local ring if R has precisely one maximal ideal. This is equivalent to saying the nonunits in R form an ideal. Thus, if R is a local ring with unique maximal ideal m, then $m = U(R)^c$. In this case, $J(R) = m$. We will adopt the notation (R, m, k) to denote a local ring R with maximal ideal m and residue class field $k = R/m$.

Local rings are very important in commutative ring theory as well as algebraic geometry. Let us consider some examples. Clearly, any field F is a local ring

with maximal ideal (0) and residue class field F (i.e., $(F, (0), F)$ is a local ring). If $X_1, \ldots, X_n$ are indeterminates over F, then $F[[X_1, \ldots, X_n]]$, the ring of formal power series, is a local ring with maximal ideal $(X_1, \ldots, X_n)$ and residue class field F. See Exercise 4 at the end of this chapter for more details. The p-adic integers $\hat{\mathbb{Z}}_p$ (see Exercise 11 of Chapter 1) are a local ring with maximal ideal (p) and residue class field $\mathbb{Z}/p\mathbb{Z}$.

If R is any commutative ring and $\mathfrak{P}$ is a prime ideal of R, then we can construct a local ring $(R_{\mathfrak{P}}, \mathfrak{P}R_{\mathfrak{P}}, Q(R/\mathfrak{P}))$ along the same general lines as Example 6.7. This construction is important enough to warrant a careful explanation.

Example 10.17 Let R be a commutative ring and $\mathfrak{P}$ a prime ideal of R. Consider the set $R \times \mathfrak{P}^c = \{(x, y) \mid x \in R, y \in R \cap \mathfrak{P}^c\}$. Define a relation $\sim$ on $R \times \mathfrak{P}^c$ as follows:

10.18 $(x, y) \sim (x', y')$ if $t(xy' - x'y) = 0$ for some $t \in \mathfrak{P}^c$.

It is easy to check that $\sim$ is an equivalence relation on $R \times \mathfrak{P}^c$. We will let x/y denote the equivalence class of (x, y) in $R \times \mathfrak{P}^c$. Set $R_{\mathfrak{P}} = \{x/y \mid (x, y) \in R \times \mathfrak{P}^c\}$. Notice then that $x/y = x'/y'$ in $R_{\mathfrak{P}}$ if and only if $t(xy' - x'y) = 0$ for some $t \in \mathfrak{P}^c$. Addition and multiplication on $R_{\mathfrak{P}}$ are defined by the following equations:

10.19
$$x/y + x'/y' = (xy' + x'y)/yy'$$
$$(x/y)(x'/y') = xx'/yy'$$

The reader can easily check that both of these operations are well defined and endow $R_{\mathfrak{P}}$ with the structure of a commutative ring. The ring $R_{\mathfrak{P}}$ is a local ring with (unique) maximal ideal $\mathfrak{P}R_{\mathfrak{P}} = \{x/y \in R_{\mathfrak{P}} \mid x \in \mathfrak{P}\}$ and residue class field $Q(R/\mathfrak{P})$ (the quotient field of the integral domain $R/\mathfrak{P}$). See Exercise 5 at the end of this chapter for more details.

The local ring $R_{\mathfrak{P}}$ is called the localization of R at $\mathfrak{P}$. ∎

Returning to Nakayama's lemma, we have the following corollary for local rings.

Corollary 10.20 Suppose M is a finitely generated module over a local ring (R, m, k). A set of elements $m_1, \ldots, m_n$ in M is a basis of M if and only if the images of these elements in the k-vector space M/mM span M/mM.

Proof. Let $\overline{m}_1, \ldots, \overline{m}_n$ denote the images of $m_1, \ldots, m_n$, respectively, in M/mM. Suppose $\overline{m}_1, \ldots, \overline{m}_n$ span the k-vector space M/mM. Thus, $\sum_{i=1}^n k\overline{m}_i = M/mM$. Back in M, this implies $M = (\sum_{i=1}^n Rm_i) + mM$. In particular, Corollary 10.16 implies $M = \sum_{i=1}^n Rm_i$. Thus, $\{m_1, \ldots, m_n\}$ is a basis of M.

If $\{m_1, \ldots, m_n\}$ is a basis of M, then clearly $\{\overline{m}_1, \ldots, \overline{m}_n\}$ span the k-vector space M/mM. ∎

Notice that Corollary 10.20 implies the minimum number of generators of a finitely generated module M over a local ring (R, m, k) is precisely the vector space dimension $(\dim_k(M/mM))$ of M/mM over the field k. Another application of Corollary 10.20 is the following result.

Corollary 10.21 Let $\mathfrak{A}$ be a finitely generated ideal in a local ring (R, m, k). Suppose r is the smallest number of generators of $\mathfrak{A}$. Then any set of generators of $\mathfrak{A}$ contains a set of r generators of $\mathfrak{A}$.

Proof. Since r is the minimum number of generators of the R-module $\mathfrak{A}$, our remarks before this corollary imply $\dim_k(\mathfrak{A}/m\mathfrak{A}) = r$. Suppose $\{x_1, \ldots, x_n\}$ is a set of generators of $\mathfrak{A}$. Thus, $\mathfrak{A} = \sum_{i=1}^{n} Rx_i$. Let $\bar{x}_1, \ldots, \bar{x}_n$ denote the images of $x_1, \ldots, x_n$ respectively in $\mathfrak{A}/m\mathfrak{A}$. Since $x_1, \ldots, x_n$ generate the R-module $\mathfrak{A}$, $\bar{x}_1, \ldots, \bar{x}_n$ span the k-vector space $\mathfrak{A}/m\mathfrak{A}$. In particular, $\{\bar{x}_1, \ldots, \bar{x}_n\}$ contains a k-vector space basis of $\mathfrak{A}/m\mathfrak{A}$. After suitably relabeling if need be, we can assume $\{\bar{x}_1, \ldots, \bar{x}_r\}$ is a k-vector space basis of $\mathfrak{A}/m\mathfrak{A}$. Corollary 10.20 then implies $\mathfrak{A} = \sum_{i=1}^{r} Rx_i$. Thus, $\{x_1, \ldots, x_r\}$ is a set of generators of $\mathfrak{A}$. ∎

We will need one last application of Nakayama's lemma in Chapter 13. An R-module M is called a direct summand of a free R-module if there exist a free R-module F containing M and a submodule N of F such that $M \oplus N = F$. If M is a direct summand of the free R-module F and F is a finitely generated R-module, then M is also finitely generated. For if $M \oplus N = F$, then $M \cong F/N$ a finitely generated R-module. Our last corollary in this chapter says direct summands of R^n are free R-modules when R is local.

Corollary 10.22 Let (R, m, k) be a local ring. If M is a direct summand of a finitely generated, free R-module F, then M is a free R-module.

Proof. M is a direct summand of a finitely generated, free R-module. Replacing M with an isomorphic copy if need be, we can assume there exists a positive integer n such that $M \subseteq R^n$ and $M \oplus N = R^n$ for some submodule $N \subseteq R^n$. Then $mR^n = mM \oplus mN$, and

$$R^n/mR^n = (M + mR^n/mR^n) \oplus (N + mR^n/mR^n)$$

As usual, let $\mathfrak{e} = \{\varepsilon_1, \ldots, \varepsilon_n\}$ denote the canonical basis of R^n. It follows from Theorem 5.10 that any set of generators of R^n must contain at least n elements. In particular, Corollary 10.20 implies $\dim_k(R^n/mR^n) = n$.

Since $R^n/mR^n = (M + mR^n/mR^n) \oplus (N + mR^n)/mR^n)$, we can select a k-vector space basis $\{\bar{y}_1, \ldots, \bar{y}_n\}$ of R^n/mR^n such that $y_1, \ldots, y_r \in M$

and $y_{r+1}, \ldots, y_n \in N$. Here $r = \dim_k((M + mR^n)/mR^n)$ and $n - r = \dim_k((N + mR^n)/mR^n)$. Again Corollary 10.20 implies $\{y_1, \ldots, y_n\}$ is an R-module basis of R^n. It then follows from Theorem 5.10 that $\{y_1, \ldots, y_n\}$ is a free R-module basis of R^n. Since $\{y_1, \ldots, y_r\} \subseteq M$ and $\{y_{r+1}, \ldots, y_n\} \subseteq N$, it is easy to check that $\{y_1, \ldots, y_r\}$ is a free R-module basis of M. ∎

EXERCISES

1. Let $\mathcal{F}$ be a nonempty set of submodules of the R-module M. Let $\mathcal{S}$ denote the set of all strictly ascending chains $N_1 < N_2 < N_3 < \cdots$ (finite or infinite) of submodules from $\mathcal{F}$. Partially order ($\lll$) the chains in $\mathcal{S}$ as follows: If $A = \{N_1 < N_2 < \cdots < N_r\}$ and $B = \{M_1 < M_2 < \cdots < M_s\}$ are finite chains in $\mathcal{S}$, then set $A \lll B$ if $r \leq s$ and $N_i = M_i$ for all $i = 1, \ldots, r$. If A is finite and $C = \{P_1 < P_2 < \cdots\}$ is infinite, set $A \lll C$ if $N_i = P_i$ for all $i = 1, \ldots, r$. Finally, if $D = \{Q_1 < Q_2 < \cdots\}$ is also infinite, set $C \lll D$ if $P_i = Q_i$ for all $i \geq 1$.
   (a) Show $\lll$ is a partial ordering on $\mathcal{S}$.
   (b) Show every totally ordered subset of $(\mathcal{S}, \lll)$ has an upper bound in $\mathcal{S}$.
   (c) Suppose $\mathcal{F}$ contains no maximal element with respect to inclusion. Use Zorn's lemma on $(\mathcal{S}, \lll)$ to argue there exists an infinite, strictly ascending chain $N_1 < N_2 < N_3 \cdots$ in $\mathcal{S}$.
2. Show the ideal $\mathfrak{A} = (X_1, X_2, \ldots)$ in Example 10.6 is not finitely generated.
3. Prove the assertion in equation 10.14.
4. Let F be a field and $X_1, \ldots, X_n$ indeterminates over F. The power series ring $F[[X_1, \ldots, X_n]]$ is defined inductively as $F[[X_1, \ldots, X_n]] = (F[[X_1, \ldots, X_{n-1}]])[[X_n]]$. Show $F[[X_1, \ldots, X_n]]$ is a local ring with maximal ideal $(X_1, \ldots, X_n)$ and residue class field F. $F[[X_1, \ldots, X_n]]$ is a Noetherian ring also. See [1].)
5. Complete all the details of Example 10.17:
   (a) Show $\sim$ is an equivalence relation on $R \times \mathfrak{P}^c$.
   (b) Show that the binary operations defined in equation 10.19 are well defined.
   (c) Show $R_{\mathfrak{P}}$ is a commutative ring with these operations.
   (d) Show $R_{\mathfrak{P}}$ is a local ring with maximal ideal $\mathfrak{P}R_{\mathfrak{P}}$ and residue class field $Q(R/\mathfrak{P})$.
6. Suppose M is a finitely generated module over the local ring (R, m, k). Show $\dim_k(M/mM)$ is the smallest number of generators of M.
7. Give an example which shows that the statement in Corollary 10.21 is not true if R is not local.
8. Give an example of an R-module which is not a direct summand of a free module.
9. Show that a direct summand of a finitely generated, free R-module need not be free if R is not local.

10. Complete the argument in the proof of Corollary 10.22 by showing $\{y_1, \ldots, y_r\}$ is a free R-module basis of M.
11. Suppose R and S are commutative rings such that S is a finitely generated R-module. Show any finitely generated S-module is a finitely generated R-module.
12. Suppose M and N are R-modules with N finitely generated. Let $f \in \mathrm{Hom}_R(M, N)$. Let $\mathfrak{A} \subseteq J(R)$. If the induced mapping $\bar{f} : M/\mathfrak{A}M \mapsto N/\mathfrak{A}N$ is surjective, show f is surjective.
13. Is Exercise 12 true with the word surjective replaced by the word injective?
14. Suppose M is a finitely generated R-module. Let $f \in \mathrm{Hom}_R(M, R^n)$. If f is surjective, show $\mathrm{Ker}(f)$ is a finitely generated R-module.
15. Suppose M is a finitely presented R-module. If

$$(0) \mapsto N \overset{g}{\mapsto} P \overset{f}{\mapsto} M \mapsto (0)$$

is a short exact sequence of R-modules and P is finitely generated, show N is finitely generated.
16. Use the results in Exercise 15 to show $R/\mathfrak{A}$ in Example 10.6 is not finitely presented.
17. Let $R = F[X, Y]$. Set $\mathfrak{A} = (X, Y)$. Compute a relations matrix for the R-module $\mathfrak{A}$. Can you generalize your result to n variables?
18. Compute relations matrices for the following $\mathbb{Z}$-modules:
    (a) $M = \mathbb{Z}/n\mathbb{Z}$
    (b) $M = \mathbb{Z}/n\mathbb{Z} \oplus \mathbb{Z}/m\mathbb{Z}$
    (c) $M = \mathbb{Z}/n\mathbb{Z} \oplus \mathbb{Z}/m\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z}$

11

Primary Decompositions In Noetherian Rings

In this chapter, we describe primary decompositions of ideals in Noetherian rings. There are many good references for this material. Our treatment of this subject will basically be the same as that in [1] and [10]. The reader can consult these references for more details.

As usual, R will denote a commutative ring. We begin with the definition of a primary ideal of R. As we will see, primary ideals are the fundamental building blocks of all ideals in a Noetherian ring.

Definition 11.1 Let $\mathfrak{Q}$ be an ideal of R such that $\mathfrak{Q} \neq R$. The ideal $\mathfrak{Q}$ is called a primary ideal if, whenever $xy \in \mathfrak{Q}$, and $x \notin \mathfrak{Q}$, then $y^n \in \mathfrak{Q}$ for some positive integer n.

We already have seen many examples of primary ideals in this text. For example, any prime ideal of R is a primary ideal. On the other hand, there are many primary ideals which are not prime. For example, if $\mathfrak{M}$ is a maximal ideal of R, then $\mathfrak{M}^n$ is a primary ideal for all $n > 0$. This follows from Theorem 11.3 below. Notice that the improper ideal R is not called a primary ideal in this text. Thus, if $\mathfrak{Q}$ is a primary ideal of R, then $\mathfrak{Q} < R$. On the other hand, the zero ideal (0) may very well be a primary ideal of R. For example, if X and Y are indeterminates over the field $\mathbb{Q}$, then (0) is a primary ideal of the ring $\mathbb{Q}[X]/(X^2)$, but (0) is not a primary ideal of the ring $\mathbb{Q}[X,Y]/(XY)$.

We can restate the definition of $\mathfrak{Q}$ being a primary ideal of R in terms of the quotient ring $R/\mathfrak{Q}$ as follows:

11.2 A proper ideal $\mathfrak{Q}$ is a primary ideal of R if and only if every zero divisor in $R/\mathfrak{Q}$ is nilpotent.

For example, let X and Y denote indeterminates over $\mathbb{Q}$, and let $\mathfrak{Q} = (X,Y^2) \subsetneq \mathbb{Q}[X,Y]$. $\mathfrak{Q}$ is a primary ideal of $\mathbb{Q}[X,Y]$ since every zero divisor in $\mathbb{Q}[X,Y]/\mathfrak{Q} \cong \mathbb{Q}[Y]/(Y^2)$ is obviously nilpotent. Notice that $\mathfrak{Q}$ is not any power of any maximal ideal in $\mathbb{Q}[X,Y]$.

Theorem 11.3 Let $\mathfrak{Q}$ be a proper ideal of R. Set $\mathfrak{P} = \sqrt{\mathfrak{Q}}$.

(a) If $\mathfrak{Q}$ is primary, then $\mathfrak{P}$ is a prime ideal of R.

(b) If $\mathfrak{P}$ is a maximal ideal of R, then $\mathfrak{Q}$ is a primary ideal.

Proof. (a) Suppose $xy \in \mathfrak{P}$. Then $(xy)^n \in \mathfrak{Q}$ for some positive integer n. If $x^n \in \mathfrak{Q}$, then $x \in \mathfrak{P}$. Suppose $x^n \notin \mathfrak{Q}$. Since $\mathfrak{Q}$ is primary, $(y^n)^m \in \mathfrak{Q}$ for some positive integer m. Thus, $y^{mn} \in \mathfrak{Q}$. In particular, $y \in \sqrt{\mathfrak{Q}} = \mathfrak{P}$. Thus, $\mathfrak{P}$ is a prime ideal of R.

(b) Suppose $xy \in \mathfrak{Q}$ and $y \notin \mathfrak{P}$. We must argue $x \in \mathfrak{Q}$. Since $\mathfrak{P}$ is a maximal ideal of R, $\mathfrak{P} + Ry = R$. Therefore, $1 = c + dy$ for some $c \in \mathfrak{P}$ and $d \in R$. Since $\mathfrak{P} = \sqrt{\mathfrak{Q}}$, $c^n \in \mathfrak{Q}$ for some $n > 0$. Thus, $1 = 1^n = (c + dy)^n = c^n + d'y$ for some $d' \in R$. In particular, $x = x1 = xc^n + (xy)d' \in \mathfrak{Q}$. ∎

If $\mathfrak{Q}$ is a primary ideal of R, and $\mathfrak{P} = \sqrt{\mathfrak{Q}}$, then $\mathfrak{Q}$ is called a $\mathfrak{P}$-primary ideal. For example, in $R = \mathbb{Q}[X,Y]$, $\sqrt{(X,Y^2)} = (X,Y) = \mathfrak{P}$ is a maximal ideal of R. Thus, (X,Y^2) is a $\mathfrak{P}$-primary ideal. Similarly, (0) is an (x)-primary ideal in $R = \mathbb{Q}[X]/(X^2)$. Here x denotes the image of X in R.

If $\mathfrak{P}$ is a maximal ideal of R, then Theorem 11.3b implies $\mathfrak{P}^n$ is a $\mathfrak{P}$-primary ideal for all positive integers n. However, a $\mathfrak{P}$-primary ideal $\mathfrak{Q}$ need be no power of $\mathfrak{P}$, as the first example in the last paragraph indicates.

If $\mathfrak{P}$ is a maximal ideal of R, then we have observed that $\mathfrak{P}^n$ is a $\mathfrak{P}$-primary ideal. If $\mathfrak{P}$ is an arbitrary prime of R (not necessarily maximal), then $\mathfrak{P}^n$ need not be a primary ideal. Consider the following well-known example:

Example 11.4 Let X, Y, and Z be indeterminates over the field F. Set $R = F[X,Y,Z]/(XY - Z^2)$. Let x, y, and z denote the images of X, Y, and Z, respectively, in R. Then $R = F[x,y,z]$. Let $\mathfrak{P} = (x,z)$. Since $R/\mathfrak{P} \cong F[Y]$, $\mathfrak{P}$ is a prime ideal of R. In R, $xy = z^2 \in \mathfrak{P}^2$. The reader can

easily check that $x \notin \mathfrak{P}^2$ and $y \notin \mathfrak{P} = \sqrt{\mathfrak{P}^2}$. Thus, $\mathfrak{P}^2$ is not a primary ideal of R. ∎

We need a definition used in Exercise 6 of Chapter 9.

Definition 11.5 Let $\mathfrak{A}$ and $\mathfrak{B}$ be two ideals of R. Then

$$\mathfrak{A} : \mathfrak{B} = \{x \in R \mid x\mathfrak{B} \subseteq \mathfrak{A}\}.$$

It is easy to check that $\mathfrak{A} : \mathfrak{B}$ is an ideal of R containing $\mathfrak{A}$. The ideal $\mathfrak{A} : \mathfrak{B}$ is usually called the quotient of $\mathfrak{A}$ by $\mathfrak{B}$. We will use this construction in our proof of the Lasker-Noether decomposition theorem.

Theorem 11.6 (Lasker-Noether) Let R be a Noetherian ring. Let $\mathfrak{A}$ be a proper ideal of R. Then $\mathfrak{A}$ is a finite intersection of primary ideals.

Proof. An ideal $\mathfrak{A} < R$ is said to be irreducible if $\mathfrak{A}$ is not the intersection of two ideals strictly containing it. We first claim every proper ideal of R is a finite intersection of irreducible ideals. To see this, suppose the claim is false. Let $\mathfrak{F}$ denote the set of all proper ideals of R which are not finite intersections of irreducible ideals. We are assuming the claim is false. Consequently, $\mathfrak{F}$ is a nonempty set. Since R is Noetherian, $\mathfrak{F}$ contains a maximal element by Lemma 10.2b. Let $\mathfrak{B}$ be a maximal element of $\mathfrak{F}$. Since $\mathfrak{B} \in \mathfrak{F}$, $\mathfrak{B}$ is not irreducible. Thus, $\mathfrak{B} = \mathfrak{C}_1 \cap \mathfrak{C}_2$ for two ideals $\mathfrak{C}_1$ and $\mathfrak{C}_2$ properly containing $\mathfrak{B}$. The ideals $\mathfrak{C}_1$ and $\mathfrak{C}_2$ are not in $\mathfrak{F}$ since $\mathfrak{B}$ is a maximal element of $\mathfrak{F}$ with respect to inclusion. Thus, each ideal $\mathfrak{C}_i$ is a finite intersection of irreducible ideals. But then $\mathfrak{B}$ is a finite intersection of irreducible ideals. This is impossible. We conclude that $\mathfrak{F} = \varnothing$, and every proper ideal of R is a finite intersection of irreducible ideals.

We next claim that every irreducible ideal of R is primary. To see this, suppose R contains an irreducible ideal $\mathfrak{Q}$ which is not primary. Then there exist $x, y \in \mathfrak{Q}^c$ such that $xy \in \mathfrak{Q}$ and $y \notin \mathfrak{P} = \sqrt{\mathfrak{Q}}$. Consider the chain of ideals $\mathfrak{Q} \subseteq \mathfrak{Q}: Ry \subseteq \mathfrak{Q} : Ry^2 \subseteq \mathfrak{Q} : Ry^3 \subseteq \cdots$ in R. Since R is Noetherian, there exists a positive integer n such that $\mathfrak{Q} : Ry^n = \mathfrak{Q} : Ry^{n+1}$. We then have

$$\mathfrak{Q} = (\mathfrak{Q} + Ry^n) \cap (\mathfrak{Q} + Rx) \tag{11.7}$$

Clearly, $\mathfrak{Q}$ is contained in the intersection on the right in equation 11.7. Suppose $z \in (\mathfrak{Q} + Ry^n) \cap (\mathfrak{Q} + Rx)$. Then $z = u + by^n = v + cx$ for elements u, $v \in \mathfrak{Q}$ and $b, c \in R$. Since $xy \in \mathfrak{Q}$, $yz \in \mathfrak{Q}$. Therefore, $by^{n+1} \in \mathfrak{Q}$. In particular, $b \in \mathfrak{Q} : Ry^{n+1} = \mathfrak{Q} : Ry^n$. Thus, $by^n \in \mathfrak{Q}$. But then $z \in \mathfrak{Q}$. This establishes the equality in equation 11.7.

Now x and y^n are not elements in $\mathfrak{Q}$. Therefore, the ideals $\mathfrak{Q} + Ry^n$ and $\mathfrak{Q} + Rx$ properly contain $\mathfrak{Q}$. Equation 11.7 implies $\mathfrak{Q}$ is not irreducible. This is impossible. We conclude that every irreducible ideal of R is primary.

We can now prove the theorem. Suppose $\mathfrak{A}$ is an ideal of R with $\mathfrak{A} \neq R$. By our first claim, $\mathfrak{A} = \mathfrak{Q}_1 \cap \mathfrak{Q}_2 \cap \cdots \cap \mathfrak{Q}_n$ with each $\mathfrak{Q}_i$ an irreducible ideal of R. By our second claim, each $\mathfrak{Q}_i$ is a primary ideal of R. Thus, $\mathfrak{A}$ is a finite intersection of primary ideals. ∎

We have shown in the proof of Theorem 11.6 that irreducible ideals are primary in a Noetherian ring. The converse of this statement is not true. A primary ideal need not be irreducible. For example, suppose $R = F[X,Y]$ is a polynomial ring in two indeterminates X and Y over a field F. Then $\mathfrak{Q} = (X^2,XY,Y^2) = (X,Y)^2$ is a primary ideal by Theorem 11.3b. However, $\mathfrak{Q}$ is not irreducible since $\mathfrak{Q} = (X,Y^2) \cap (X^2,Y)$.

There is a sharper version of Theorem 11.6 which we list as the following corollary.

Corollary 11.8 Let $\mathfrak{A}$ be a proper ideal in a Noetherian ring R. Then there exist primary ideals $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_n$ of R such that the following conditions are satisfied:

(a) $\mathfrak{A} = \mathfrak{Q}_1 \cap \mathfrak{Q}_2 \cap \cdots \cap \mathfrak{Q}_n$.

(b) For each $i = 1, \ldots, n$

$$\mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_{i-1} \cap \mathfrak{Q}_{i+1} \cap \cdots \cap \mathfrak{Q}_n$$

is not contained in $\mathfrak{Q}_i$.

(c) $\sqrt{\mathfrak{Q}_1}, \ldots, \sqrt{\mathfrak{Q}_n}$ are distinct (primes) of R.

Proof. Theorem 11.6 implies $\mathfrak{A} = \mathfrak{Q}_1 \cap \mathfrak{Q}_2 \cap \cdots \cap \mathfrak{Q}_n$ for some primary ideals $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_n$ of R. If some $\mathfrak{Q}_i$ contains the intersection of the remaining $\mathfrak{Q}_j$, then we can drop $\mathfrak{Q}_i$ from $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_n$ and write

$$\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_{i-1} \cap \mathfrak{Q}_{i+1} \cap \cdots \cap \mathfrak{Q}_n$$

Hence, we can always find primary ideals $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_n$ such that (a) and (b) are satisfied.

We have seen in Theorem 11.3 that $\sqrt{\mathfrak{Q}_i} = \mathfrak{P}_i$ is a prime ideal of R. Suppose $\sqrt{\mathfrak{Q}_1} = \mathfrak{P}_1 = \sqrt{\mathfrak{Q}_2}$. Then $\mathfrak{Q}_1 \cap \mathfrak{Q}_2$ is a $\mathfrak{P}_1$-primary ideal. To see this, first observe that

$$\sqrt{\mathfrak{Q}_1 \cap \mathfrak{Q}_2} = \sqrt{\mathfrak{Q}_1} \cap \sqrt{\mathfrak{Q}_2} = \mathfrak{P}_1 \cap \mathfrak{P}_1 = \mathfrak{P}_1$$

Thus, $\mathfrak{P}_1$ is the radical of $\mathfrak{Q}_1 \cap \mathfrak{Q}_2$. Suppose $xy \in \mathfrak{Q}_1 \cap \mathfrak{Q}_2$ and $y \notin \mathfrak{P}_1$. Since $\mathfrak{Q}_1$ is $\mathfrak{P}_1$-primary, we conclude $x \in \mathfrak{Q}_1$. Similarly, $x \in \mathfrak{Q}_2$. Thus, $\mathfrak{Q}_1 \cap \mathfrak{Q}_2$ is

$\mathfrak{P}_1$-primary. By induction, any finite intersection of $\mathfrak{P}_1$-primary ideals is a $\mathfrak{P}_1$-primary ideal.

Now suppose $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_n$ satisfy conditions (a) and (b) in Corollary 11.8. Suppose we have labeled that $\mathfrak{Q}_i$ such that $\mathfrak{P}_1, \ldots, \mathfrak{P}_r$ are the distinct primes in the list $\sqrt{\mathfrak{Q}_1}, \ldots, \sqrt{\mathfrak{Q}_n}$. Here $1 \leq r \leq n$. Fix $i \in \{1, \ldots, r\}$. Let $\mathfrak{Q}_i'$ denote the intersection of those $\mathfrak{Q}_j$ which have radical $\mathfrak{P}_i$. From the last paragraph, $\mathfrak{Q}_i'$ is a $\mathfrak{P}_i$-primary ideal. Clearly, $\mathfrak{A} = \mathfrak{Q}_1' \cap \cdots \cap \mathfrak{Q}_r'$, and $\mathfrak{Q}_1', \ldots, \mathfrak{Q}_r'$ satisfy conditions (a), (b), and (c). ∎

Any primary decomposition $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ in which $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_n$ are primary ideals satisfying conditions (a), (b), and (c) in Corollary 11.8 is called an irredundant, primary decomposition of $\mathfrak{A}$. Corollary 11.8 guarantees every proper ideal $\mathfrak{A}$ in a Noetherian ring has an irredundant, primary decomposition. The prime ideals $\mathfrak{P}_1 = \sqrt{\mathfrak{Q}_1}, \ldots, \mathfrak{P}_n = \sqrt{\mathfrak{Q}_n}$ in an irredundant, primary decomposition of $\mathfrak{A}$ are called the associated primes of $\mathfrak{A}$. As we will soon see, the associated primes of $\mathfrak{A}$ are unique. By this we mean that any two irredundant, primary decompositions of $\mathfrak{A}$ have the same set of associated primes.

Before proving the uniqueness theorem, we point out that irredundant, primary decompositions themselves are not unique. In fact, a given ideal could have infinitely many different irredundant, primary decompositions. Consider the following well-known example.

Example 11.9 Let X and Y be indeterminates over the field F, and set $R = F[X,Y]$. Let $\mathfrak{A} = (X^2, XY)$. Then for any $c \in F$, $\mathfrak{A} = (X) \cap (Y - cX, X^2)$ is an irredundant, primary decomposition of $\mathfrak{A}$. For different c's, the primary ideals $(Y - cX, X^2)$ are different. Thus, if F is infinite, $\mathfrak{A}$ has infinitely many different irreudundant, primary decompositions. ∎

In Example 11.9, $\sqrt{(X)} = (X) = \mathfrak{P}_1$ and $\sqrt{(Y-cX, X^2)} = (X,Y) = \mathfrak{P}_2$. Thus, $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are associated primes of $\mathfrak{A}$. It turns out that any irredundant, primary decomposition of $\mathfrak{A}$ has only $\mathfrak{P}_1$ and $\mathfrak{P}_2$ as associated primes. This will follow from our next theorem.

Our next theorem will require two lemmas which are of some interest in their own right.

Lemma 11.10

(a) Suppose $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ and $\mathfrak{P}$ are ideals of R. If $\mathfrak{P}$ is prime and $\mathfrak{A}_1 \cap \mathfrak{A}_2 \cap \cdots \cap \mathfrak{A}_n \subseteq \mathfrak{P}$, then $\mathfrak{A}_i \subseteq \mathfrak{P}$ for some i.

(b) Suppose $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ are prime ideals of R. If $\mathfrak{A}$ is an ideal of R such that $\mathfrak{A} \subseteq \mathfrak{P}_1 \cup \cdots \cup \mathfrak{P}_n$, then $\mathfrak{A} \subseteq \mathfrak{P}_i$ for some i.

Proof. (a) Suppose no $\mathfrak{A}_i$ is contained in $\mathfrak{P}$. Then for each $i = 1, \ldots, n$, there exists an element $x_i \in \mathfrak{A}_i - \mathfrak{P}$. Since $\mathfrak{P}$ is prime, $x_1x_2 \cdots x_n \notin \mathfrak{P}$. But, $x_1x_2 \cdots x_n \in \mathfrak{A}_1\mathfrak{A}_2 \cdots \mathfrak{A}_n \subseteq \cap_{i=1}^n \mathfrak{A}_i \subseteq \mathfrak{P}$. This is impossible. We conclude that some ideal $\mathfrak{A}_i$ is contained in $\mathfrak{P}$.

(b) Suppose $\mathfrak{A} \subseteq \cup_{i=1}^n \mathfrak{P}_i$. We can assume with no loss of generality that $\mathfrak{P}_i$ is not contained in $\mathfrak{P}_j$ whenever $i \neq j$. Let us suppose $\mathfrak{A}$ is not contained in any $\mathfrak{P}_i$ and derive a contradiction.

Since there are no containment relations among the $\mathfrak{P}_i$ and $\mathfrak{A}$ is not contained in any $\mathfrak{P}_i$, $\mathfrak{A} \cap (\cap_{j \neq i} \mathfrak{P}_j)$ is not contained in $\mathfrak{P}_i$ for each $i = 1, \ldots, n$. This statement follows from (a). Let $a_i \in \mathfrak{A} \cap (\cap_{j \neq i} \mathfrak{P}_j) \cap \mathfrak{P}_i^c$. Then $a_1 + \cdots + a_n$ is in $\mathfrak{A}$ but not in any $\mathfrak{P}_i$ $(i = 1, \ldots, n)$. For suppose $a_1 + \cdots + a_n \in \mathfrak{P}_k$. Since $a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_n \in \mathfrak{P}_k$,

$$a_k = (a_1 + \cdots + a_n) - (a_1 + \cdots + a_{k-1} + a_{k+1} + \cdots + a_n) \in \mathfrak{P}_k$$

This is impossible. Since $a_1 + \cdots + a_n \in \mathfrak{A} \subseteq \cup_{i=1}^n \mathfrak{P}_i$, we have a contradiction. ∎

Lemma 11.11 Let $\mathfrak{Q}$ be a $\mathfrak{P}$-primary ideal of R. Let $x \in R$.

(a) If $x \in \mathfrak{Q}$, then $\mathfrak{Q} : Rx = R$.

(b) If $x \notin \mathfrak{Q}$, then $\mathfrak{Q} : Rx$ is a $\mathfrak{P}$-primary ideal.

(c) If $x \notin \mathfrak{P}$, then $\mathfrak{Q} : Rx = \mathfrak{Q}$.

Proof. (a) and (c) follow easily from the definitions. We prove (b). If $y \in \mathfrak{Q} : Rx$, then $xy \in \mathfrak{Q}$. Since $x \notin \mathfrak{Q}$, $y \in \mathfrak{P}$. Therefore, $\mathfrak{Q} \subseteq \mathfrak{Q} : Rx \subseteq \mathfrak{P} = \sqrt{\mathfrak{Q}}$. We conclude from this that $\sqrt{\mathfrak{Q} : Rx} = \mathfrak{P}$.

Now suppose $yz \in \mathfrak{Q} : Rx$ and $y \notin \mathfrak{P}$. Then $xyz \in \mathfrak{Q}$. Since $y \notin \mathfrak{P}$, $xz \in \mathfrak{Q}$. Thus, $z \in \mathfrak{Q} : Rx$. Thus, $\mathfrak{Q} : Rx$ is a $\mathfrak{P}$-primary ideal. ∎

We can now argue that the associated primes of an ideal are unique. This result does not require that R be Noetherian. All we need is an irredundant, primary decomposition of the ideal.

Theorem 11.12 Suppose $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ is an irredundant, primary decomposition of $\mathfrak{A}$ in R. Let $\mathfrak{P}_1 = \sqrt{\mathfrak{Q}_1}, \ldots, \mathfrak{P}_n = \sqrt{\mathfrak{Q}_n}$ be the (distinct) associated primes of this decomposition. Then $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ are precisely the prime ideals in the set $\{\sqrt{\mathfrak{A} : Rx} \mid x \in R\}$.

Before proving Theorem 11.12, let us make a few remarks about this result. For every $x \in R$, $\sqrt{\mathfrak{A} : Rx}$ is an ideal of R. Hence $\Gamma = \{\sqrt{\mathfrak{A} : Rx} \mid x \in R\}$ is a (nonempty) set of ideals of R. Theorem 11.12 says those ideals in Γ which are

prime ideals are precisely $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$. Notice that this characterization of $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ does not depend on any particular irredundant, primary decomposition of $\mathfrak{A}$. Hence, if $\mathfrak{A}$ has another irredundant, primary decomposition $\mathfrak{A} = \mathfrak{Q}_1' \cap \cdots \cap \mathfrak{Q}_m'$ with associated primes $\mathfrak{P}_1' = \sqrt{\mathfrak{Q}_1'}, \ldots, \mathfrak{P}_m' = \sqrt{\mathfrak{Q}_m'}$, then $m = n$ and $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\} = \{\mathfrak{P}_1', \ldots, \mathfrak{P}_n'\}$. This is what is meant by the expression "the associated primes of $\mathfrak{A}$ are unique."

Proof of Theorem 11.12. Let $x \in R$. Then $\mathfrak{A} : Rx = (\cap_{i=1}^n \mathfrak{Q}_i) : Rx = \cap_{i=1}^n (\mathfrak{Q}_i : Rx)$. If $x \notin \mathfrak{A}$, then Lemma 11.11 implies $\mathfrak{A} : Rx = \cap \{\mathfrak{Q}_j : Rx \mid x \notin \mathfrak{Q}_j\}$ and $\sqrt{\mathfrak{A} : Rx} = \cap \{\mathfrak{P}_j \mid x \notin \mathfrak{Q}_j\}$.

Now suppose $\sqrt{\mathfrak{A} : Rx}$ is a prime ideal of R. By Lemma 11.10a, $\sqrt{\mathfrak{A} : Rx} \supseteq \mathfrak{P}_j$ for some j such that $x \notin \mathfrak{Q}_j$. But $\sqrt{\mathfrak{A} : Rx} \subseteq \mathfrak{P}_j$, and consequently, $\sqrt{\mathfrak{A} : Rx} = \mathfrak{P}_j$. Therefore, any prime ideal in the set $\Gamma = \{\sqrt{\mathfrak{A} : Rx} \mid x \in R\}$ is contained in the set $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$.

Conversely, fix $i \in \{1, \ldots, n\}$. Since $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ is an irredundant, primary decomposition, $\mathfrak{Q}_i$ does not contain $\cap_{j \neq i} \mathfrak{Q}_j$. Let $z \in (\cap_{j \neq i} \mathfrak{Q}_j) \cap \mathfrak{Q}_i^c$. Our remarks in the first paragraph of this proof then imply $\sqrt{\mathfrak{A} : Rz} = \mathfrak{P}_i$. Thus, $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ are prime ideals in Γ. Therefore, the prime ideals in $\{\sqrt{\mathfrak{A} : Rx} \mid x \in R\}$ are precisely $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$. ∎

It might be a good idea to summarize what has been said about irredundant primary decompositions in Noetherian rings. Suppose R is a Noetherian ring, and let $\mathfrak{A}$ be a proper ideal of R. Corollary 11.8 implies $\mathfrak{A}$ has an irredundant primary decomposition. Thus, there exist primary ideals $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_n$ of R such that $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$, and the ideals $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_n$ satisfy conditions (b) and (c) of Corollary 11.8. The prime ideals $\mathfrak{P}_1 = \sqrt{\mathfrak{Q}_1}, \ldots, \mathfrak{P}_n = \sqrt{\mathfrak{Q}_n}$ are called the associated primes of $\mathfrak{A}$. These primes are unique in the sense that any other irredundant, primary decomposition of $\mathfrak{A}$ has the same set $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$ of associated primes.

The next order of business is to identify the minimal primes of $\mathfrak{A}$ and the maximal primes belonging to $\mathfrak{A}$ when $\mathfrak{A}$ has an irredundant, primary decomposition. We begin with the following theorem.

Theorem 11.13 Suppose $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ is an irredundant, primary decomposition of $\mathfrak{A}$. Let $\mathfrak{P}_1 = \sqrt{\mathfrak{Q}_1}, \ldots, \mathfrak{P}_n = \sqrt{\mathfrak{Q}_n}$ be the associated primes of $\mathfrak{A}$. Then $Z(R/\mathfrak{A}) = \cup_{i=1}^n \mathfrak{P}_i$.

Proof. We view $R/\mathfrak{A}$ as an R-module. Then $x \in Z(R/\mathfrak{A})$ if and only if there exists an element $b \in R$ such that $b \notin \mathfrak{A}$, and $xb \in \mathfrak{A}$. Thus, $x \in Z(R/\mathfrak{A})$ if and

only if $(\mathfrak{A} : Rx) \cap \mathfrak{A}^c \neq \varnothing$. Since $\mathfrak{A} \subseteq \mathfrak{A} : Rx$, this last inequality is equivalent to $\mathfrak{A} : Rx \neq \mathfrak{A}$. Thus, we have reduced the theorem to the following assertion:

11.14 $\{x \in R \mid \mathfrak{A} : Rx \neq \mathfrak{A}\} = \cup_{i=1}^{n} \mathfrak{P}_i$.

In proving 11.14, we can assume $\mathfrak{A} = (0)$. To see this, consider the natural homomorphism $R \mapsto R/\mathfrak{A}$ given by $z \mapsto z + \mathfrak{A}$. Set $R/\mathfrak{A} = \bar{R}$, and let $\bar{z} = z + \mathfrak{A}$ in $\bar{R}$. If $\bar{\mathfrak{Q}}_i = \{\bar{z} \in \bar{R} \mid z \in \mathfrak{Q}_i\}$, then it is easy to check that $(\bar{0}) = \bar{\mathfrak{Q}}_1 \cap \cdots \cap \bar{\mathfrak{Q}}_n$ is an irredundant, primary decomposition of $(\bar{0})$ in $\bar{R}$. The associated primes of this decomposition are $\bar{\mathfrak{P}}_i = \{\bar{z} \mid z \in \mathfrak{P}_i\}$, $i = 1, \ldots, n$. Suppose 11.14 holds for $(\bar{0})$ in $\bar{R}$. Then $\{\bar{x} \in \bar{R} \mid (\bar{0}) : \bar{R}\bar{x} \neq (\bar{0})\} = \cup_{i=1}^{n} \bar{\mathfrak{P}}_i$. Pulling this relation back to R, we easily get 11.14 for $\mathfrak{A}$ in R.

We have now reduced the proof of Theorem 11.13 to the following statement:

11.15 $\{x \in R \mid (0) : Rx \neq (0)\} = \cup_{i=1}^{n} \mathfrak{P}_i$.

In equation 11.15, we are assuming (0) has an irredundant, primary decomposition $(0) = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ with associated primes $\mathfrak{P}_1 = \sqrt{\mathfrak{Q}_1}, \ldots, \mathfrak{P}_n = \sqrt{\mathfrak{Q}_n}$. Notice that $\{x \in R \mid (0) : Rx \neq (0)\} = Z(R)$. Thus, we have reduced the theorem to the assertion that the zero divisors of R are precisely the union of the associated primes of (0) [when (0) has an irredundant primary decomposition].

We next observe that $Z(R) = \cup_{x \neq 0} \sqrt{(0) : Rx}$. We have $Z(R) = \cup_{x \neq 0} \operatorname{Ann}_R(x) = \cup_{x \neq 0} (0) : Rx$ by equation 1.13. In particular, $Z(R) \subseteq \cup_{x \neq 0} \sqrt{(0) : Rx}$. Conversely, suppose $y \in \sqrt{(0) : Rx}$ for some $x \neq 0$. Then $y^n x = 0$ for some positive integer n. We can suppose n is as small as possible here. If $n = 1$, then $y \in (0) : Rx \subseteq Z(R)$. If $n > 1$, then $y^{n-1}x \neq 0$ and $y \in (0) : R(y^{n-1}x) \subseteq Z(R)$. In either case, $y \in Z(R)$. Thus, $\cup_{x \neq 0} \sqrt{(0) : Rx} \subseteq Z(R)$.

Let $x \in R^*$. From the proof of Theorem 11.12, we have $\sqrt{(0) : Rx} \subseteq \cap \{\mathfrak{P}_j \mid x \notin \mathfrak{Q}_j\} \subseteq \mathfrak{P}_j$ for some j. Therefore, $Z(R) \subseteq \cup_{i=1}^{n} \mathfrak{P}_i$. On the other hand, we have seen in the proof of Theorem 11.12 that each $\mathfrak{P}_i$ has the form $\mathfrak{P}_i = \sqrt{(0) : Rx}$ for some $x \in R^*$. Therefore, $\cup_{i=1}^{n} \mathfrak{P}_i \subseteq Z(R)$. ∎

In Chapter 13, we will need a slightly different version of Theorem 11.13, which we present here.

Theorem 11.16 Let R be a Noetherian ring. Let $\mathfrak{A}$ and $\mathfrak{B}$ be ideals of R with $\mathfrak{A} \neq R$. Then $\mathfrak{A} : \mathfrak{B} = \mathfrak{A}$ if and only if $\mathfrak{B}$ is contained in no associated prime of $\mathfrak{A}$.

Proof. Let $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ be an irredundant, primary decomposition of $\mathfrak{A}$ with associated primes $\mathfrak{P}_1 = \sqrt{\mathfrak{Q}_1}, \ldots, \mathfrak{P}_n = \sqrt{\mathfrak{Q}_n}$. Suppose $\mathfrak{B}$ is contained in no $\mathfrak{P}_i$. Since $(\mathfrak{A} : \mathfrak{B})\mathfrak{B} \subseteq \mathfrak{A} \subseteq \mathfrak{Q}_i$ and $\mathfrak{B}$ is not contained in $\mathfrak{P}_i$, $\mathfrak{A} : \mathfrak{B} \subseteq \mathfrak{Q}_i$. Therefore, $\mathfrak{A} : \mathfrak{B} \subseteq \cap_{i=1}^n \mathfrak{Q}_i = \mathfrak{A}$. Since $\mathfrak{A} \subseteq \mathfrak{A} : \mathfrak{B}$, we have $\mathfrak{A} : \mathfrak{B} = \mathfrak{A}$.

On the other hand, suppose $\mathfrak{A} : \mathfrak{B} = \mathfrak{A}$. Then $\mathfrak{A} : \mathfrak{B}^2 = (\mathfrak{A} : \mathfrak{B}) : \mathfrak{B} = \mathfrak{A} : \mathfrak{B} = \mathfrak{A}$. By induction, we have $\mathfrak{A} : \mathfrak{B}^p = \mathfrak{A}$ for all $p \geq 1$. Suppose $\mathfrak{B}$ is contained in some $\mathfrak{P}_i$. Relabeling if need be, we can assume $\mathfrak{B} \subseteq \mathfrak{P}_1$. Since R is a Noetherian ring, $\mathfrak{B}$ is a finitely generated ideal. Since $\mathfrak{P}_1 = \sqrt{\mathfrak{Q}_1}$, there exists an integer $s > 0$ such that $\mathfrak{B}^s \subseteq \mathfrak{Q}_1$. In particular, $\mathfrak{Q}_1 : \mathfrak{B}^s = R$. But then

$$\begin{aligned}\mathfrak{A} &= \mathfrak{A} : \mathfrak{B}^s = (\cap_{i=1}^n \mathfrak{Q}_i) : \mathfrak{B}^s = \cap_{i=1}^n (\mathfrak{Q}_i : \mathfrak{B}^s) \\ &= \cap_{i=2}^n (\mathfrak{Q}_i : \mathfrak{B}^s) \supseteq \cap_{i=2}^n \mathfrak{Q}_i \supseteq \mathfrak{A}.\end{aligned}$$

Therefore, $\mathfrak{A} = \cap_{i=2}^n \mathfrak{Q}_i$. This is impossible since $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ is an irredundant, primary decomposition of $\mathfrak{A}$. We conclude that $\mathfrak{B}$ is not contained in any associated prime $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ of $\mathfrak{A}$. ∎

We can recover Theorem 11.13 from Theorem 11.16 by setting $\mathfrak{B} = Rx$. Theorem 11.16 implies x is contained in some associated prime of $\mathfrak{A}$ if and only if $xy \in \mathfrak{A}$ for some $y \notin \mathfrak{A}$. This is precisely the statement that $\cup_{i=1}^n \mathfrak{P}_i = Z(R/\mathfrak{A})$.

Now suppose R is a Noetherian ring. Let $\mathfrak{A}$ be a proper ideal of R. Then $\mathfrak{A}$ has an irredundant, primary decomposition $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ with associated primes $\mathfrak{P}_1 = \sqrt{\mathfrak{Q}_1}, \ldots, \mathfrak{P}_n = \sqrt{\mathfrak{Q}_n}$. We partially order the set of primes $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$ by inclusion $\subseteq$.

Suppose $\mathfrak{P}$ is a minimal prime of $\mathfrak{A}$. Then $\mathfrak{P} \supseteq \mathfrak{A} = \cap_{i=1}^n \mathfrak{Q}_i$. By Lemma 11.10a, $\mathfrak{P} \supseteq \mathfrak{Q}_j$ for some $j \in \{1, \ldots, n\}$. Since $\mathfrak{P}$ is a prime ideal, $\mathfrak{P} \supseteq \sqrt{\mathfrak{Q}_j} = \mathfrak{P}_j \supseteq \mathfrak{A}$. Since $\mathfrak{P}$ is a minimal prime of $\mathfrak{A}$, $\mathfrak{P} = \mathfrak{P}_j$. Thus, any minimal prime of $\mathfrak{A}$ is a prime in $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$ which is minimal with respect to inclusion.

Conversely, suppose $\mathfrak{P}_j$ is minimal in $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$. By Corollary 6.12, $\mathfrak{P}_j$ contains a minimal prime $\mathfrak{P}$ of $\mathfrak{A}$. Our discussion in the preceding paragraph shows $\mathfrak{P} = \mathfrak{P}_k$ for some $k \in \{1, \ldots, n\}$. Since $\mathfrak{P}_j$ is minimal in $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$, $\mathfrak{P}_j = \mathfrak{P}_k = \mathfrak{P}$. Thus, every minimal prime in $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$ is a minimal prime of $\mathfrak{A}$. We have now proved the following assertion:

11.17 In a Noetherian ring R, the minimal primes of a proper ideal $\mathfrak{A}$ are finite in number and are precisely the primes in $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$ (the associated primes of $\mathfrak{A}$) which are minimal with respect to inclusion.

Now suppose $\mathfrak{P}$ is a maximal prime belonging to $\mathfrak{A}$. Since $\mathfrak{P}$ belongs to $\mathfrak{A}$, $\mathfrak{P} \subseteq Z(R/\mathfrak{A})$. By Theorem 11.13, $Z(R/\mathfrak{A}) = \cup_{i=1}^n \mathfrak{P}_i$. Lemma 11.10b then

implies $\mathfrak{P} \subseteq \mathfrak{P}_i$ for some $i \in \{1, \ldots, n\}$. But $\mathfrak{P}_i$ also belongs to $\mathfrak{A}$. Consequently, the maximality of $\mathfrak{P}$ implies $\mathfrak{P} = \mathfrak{P}_i$. Since $\mathfrak{P}$ is maximal among all ideals belonging to $\mathfrak{A}$, $\mathfrak{P} = \mathfrak{P}_i$ is certainly maximal with respect to inclusion in $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$. Thus, if $\mathfrak{P}$ is a maximal prime belonging to $\mathfrak{A}$, then $\mathfrak{P}$ is an associated prime of $\mathfrak{A}$ which is maximal with respect to inclusion in the set $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$.

Conversely, suppose $\mathfrak{P}_i$ is maximal with respect to inclusion in $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$. Then $\mathfrak{P}_i$ belongs to $\mathfrak{A}$ by Theorem 11.13. A slight variation of the proof of Theorem 6.16 shows $\mathfrak{P}_i$ is contained in a maximal prime $\mathfrak{P}$ belonging to $\mathfrak{A}$. From the preceding paragraph, $\mathfrak{P} = \mathfrak{P}_j$ for some $j \in \{1, \ldots, n\}$. Since $\mathfrak{P}_i$ is maximal with respect to inclusion in $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$, we conclude $\mathfrak{P}_i = \mathfrak{P}_j = \mathfrak{P}$. In particular, $\mathfrak{P}_i$ is a maximal prime belonging to $\mathfrak{A}$. We have now proved the following assertion:

11.18 In a Noetherian ring R, the maximal primes belonging to a proper ideal $\mathfrak{A}$ are finite in number and are precisely the primes in $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$ (the associated primes of $\mathfrak{A}$) which are maximal with respect to inclusion.

The primes in $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$ which are minimal with respect to inclusion are often called isolated primes of $\mathfrak{A}$. Thus, in a Noetherian ring, the isolated primes of $\mathfrak{A}$ are precisely the minimal primes of $\mathfrak{A}$. A prime in $\{\mathfrak{P}_1, \ldots, \mathfrak{P}_n\}$ which is not isolated is called an imbedded prime of $\mathfrak{A}$.

One obvious application of this material is to null ideals of matrices over Noetherian rings. Suppose R is a Noetherian ring. Let $A \in M_{n\times n}(R)$. The Hilbert Basis Theorem implies $R[X]$ is a Noetherian ring. Therefore, the null ideal N_A of A is a finitely generated ideal. The minimal primes of N_A as well as the maximal primes belonging to N_A are finite in number. By 11.17 and 11.18, these prime ideals can be computed from any irredundant, primary decomposition of N_A. We have seen in Corollary 7.38 that the minimal primes of N_A are precisely the same as the minimal primes of the principal ideal $(C_A(X))$. Similarly, by Corollary 9.10, the maximal primes belonging to N_A are precisely the same as the maximal primes belonging to $(C_A(X))$. Thus, the ideals N_A and $(C_A(X))$ may not have the same irredundant, primary decompositions, but they do have the same set of isolated primes and the same set of maximal primes belonging to the given ideal.

EXERCISES

1. Show that the zero ideal (0) is a primary ideal in the ring $\mathbb{Q}[X]/(X^2)$ but is not a primary ideal in $\mathbb{Q}[X,Y]/(XY)$.
2. Show that the converse of Theorem 11.3a is not true in general: If $\sqrt{\mathfrak{Q}}$ is a prime ideal, $\mathfrak{Q}$ need not be a primary ideal.

3. Let $R = \mathbb{Q}[X,Y]$. Show $\mathfrak{Q} = (X,Y^2)$ is not any power of any maximal ideal of R. Thus, primary ideals need not be powers of maximal ideals.
4. In Example 11.4, show that $x \notin \mathfrak{P}^2$ and $y \notin \mathfrak{P}$.
5. Show $(X^2,XY,Y^2) = (X,Y^2) \cap (X^2,Y)$ in $F[X,Y]$. Thus, a primary ideal need not be irreducible.
6. In Example 11.9, show $(X^2,XY) = (X) \cap (Y - cX, X^2)$ for any $c \in F$. Thus, show irredundant, primary decompositions are not unique.
7. In the proof of Theorem 11.13, show $(\bar{0}) = \overline{\mathfrak{Q}}_1 \cap \cdots \cap \overline{\mathfrak{Q}}_n$ is an irredundant, primary decomposition of $(\bar{0})$ in $R/\mathfrak{A}$ with associated primes $\overline{\mathfrak{P}}_1, \ldots, \overline{\mathfrak{P}}_n$.
8. Suppose $\mathfrak{A}$ and $\mathfrak{B}$ are ideals of R and $\mathfrak{B}$ belongs to $\mathfrak{A}$. Show $\mathfrak{B}$ is contained in a maximal prime belonging to $\mathfrak{A}$.
9. Exhibit a Noetherian ring R and an ideal $\mathfrak{A}$ ($\neq R$) in R such that $\mathfrak{A}$ has an imbedded prime.
10. Let R be a Noetherian ring. Suppose $\mathfrak{A}$ is a proper ideal of R such that $\mathfrak{A} = \sqrt{\mathfrak{A}}$. Show $\mathfrak{A}$ has no imbedded primes.
11. Let $R = F[X,Y,Z]$. Here X, Y, and Z are indeterminates over the field F. Let $\mathfrak{P}_1 = (X,Y)$ and $\mathfrak{P}_2 = (X,Z)$. Find an irredundant, primary decomposition of $\mathfrak{A} = \mathfrak{P}_1\mathfrak{P}_2$.
12. Let R be a commutative ring and X an indeterminate over R. If $\mathfrak{A}$ is an ideal of R, let $\mathfrak{A}[X]$ denote the set of polynomials in $R[X]$ whose coefficients lie in $\mathfrak{A}$. Prove the following facts about the map $\mathfrak{A} \mapsto \mathfrak{A}[X]$:
    (a) If $\mathfrak{A}$ is an ideal of R, then $\mathfrak{A}[X]$ is an ideal of $R[X]$.
    (b) If $\mathfrak{P}$ is a prime ideal of R, then $\mathfrak{P}[X]$ is a prime ideal of $R[X]$.
    (c) If $\mathfrak{Q}$ is $\mathfrak{P}$-primary in R, then $\mathfrak{Q}[X]$ is $\mathfrak{P}[X]$-primary in $R[X]$.
    (d) If $\mathfrak{A} = \cap_{i=1}^n \mathfrak{Q}_i$ is an irredundant, primary decomposition of $\mathfrak{A}$ in R, then $\mathfrak{A}[X] = \cap_{i=1}^n \mathfrak{Q}_i[X]$ is an irredundant, primary decomposition of $\mathfrak{A}[X]$ in $R[X]$.
    (e) If $\mathfrak{P}$ is a minimal prime of $\mathfrak{A}$ in R, then $\mathfrak{P}[X]$ is a minimal prime of $\mathfrak{A}[X]$ in $R[X]$.
13. Let F be a field and $X_1, \ldots, X_n$ be indeterminates over F. Set $R = F[X_1, \ldots, X_n]$. Use the results from Exercise 12 above to show the following: The ideals $\mathfrak{P}_i = (X_1, \ldots, X_i)$, $i = 1, \ldots, n$, are all prime ideals of R and their powers are all primary ideals of R.
14. Suppose R is a PID. Determine the set of all primary ideals of R.
15. Determine all primary ideals of the ring $\mathbb{Z}/n\mathbb{Z}$.
16. Let $R = \mathbb{Z}[X,Y]$ with X and Y indeterminates over $\mathbb{Z}$. Let $\mathfrak{A} = (2, XY)$. Find an irredundant, primary decomposition of $\mathfrak{A}$.
17. Find an irredundant, primary decomposition of N_A for each matrix A in Exercise 4 of Chapter 9.

18. Let $R = \mathbb{C}[X,Y,Z]$. Here X, Y, and Z are indeterminates over the field of complex numbers $\mathbb{C}$. Set $i = \sqrt{-1}$. Show

$$\begin{aligned}\mathfrak{A} &= (X^2 - Z^2, ZX - Y^2) \\ &= (X - Z, Y - Z) \cap (X - Z, Y + Z) \\ &\cap (X + Z, Y - iZ) \cap (X + Z, Y + iZ)\end{aligned}$$

is an irredundant, primary decomposition of $\mathfrak{A}$.

19. In Exercise 18, show $\mathfrak{A} = (Z^2, XZ, Y^6, ZY^3) = (Z, Y^6) \cap (X, Y^3, Z^2)$ is an irredundant, primary decomposition of $\mathfrak{A}$.
20. What are the isolated and imbedded primes of $\mathfrak{A}$ in both Exercises 18 and 19?
21. Generalize Lemma 11.10b as follows: Suppose $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ are ideals of a commutative ring R. Let S be an additive subgroup of R which is closed under multiplication (e.g., S a subring of R with or without identity, or S an ideal of R). Suppose at least $n - 2$ of the $\mathfrak{A}_i$ are prime ideals. If $S \subseteq \cup_{i=1}^{n} \mathfrak{A}_i$, then $S \subseteq \mathfrak{A}_j$ for some j.

12
Tensor Products

In this chapter, we discuss bilinear mappings and their connections with tensor products. We also give a brief description of some of the more important functorial properties of the tensor product.

Tensor products are used to solve a mapping problem in the theory of bilinear functions. Hence, we begin this section with the definition of a bilinear map. As usual, R is a commutative ring.

Definition 12.1 Let M, N, and P be R-modules. A function $\psi : M \times N \mapsto P$ is called an R-bilinear mapping if the following two conditions are satisfied:

(a) For each $m \in M$, the map $n \mapsto \psi(m, n)$ is an R-module homomorphism from N to P.

(b) For each $n \in N$, the map $m \mapsto \psi(m, n)$ is an R-module homomorphism from M to P.

Thus, if $\psi : M \times N \mapsto P$ is an R-bilinear mapping, then for every $m \in M$, $\psi(m, *) \in \mathrm{Hom}_R(N, P)$, and for every $n \in N$, $\psi(*, n) \in \mathrm{Hom}_R(M, P)$. In particular, the following equations hold for all $m, m_1, m_2 \in M$, $n, n_1, n_2 \in N$, and $r_1, r_2 \in R$:

$$\textbf{12.2} \quad \begin{aligned} \psi(m, r_1n_1 + r_2n_2) &= r_1\psi(m, n_1) + r_2\psi(m, n_2) \\ \psi(r_1m_1 + r_2m_2, n) &= r_1\psi(m_1, n) + r_2\psi(m_2, n) \end{aligned}$$

Clearly, any function $\psi : M \times N \mapsto P$ which satisfies the equations in 12.2 is an R-bilinear mapping from $M \times N$ to P.

We will let $\text{Bil}_R(M \times N, P)$ denote the set of all R-bilinear mappings of $M \times N$ to P. Since the zero map from $M \times N$ to P obviously satisfies the equations in 12.2, $\text{Bil}_R(M \times N, P) \neq \varnothing$. When the ring R is clear from the context, we will usually drop any reference to R. Thus, an R-bilinear mapping from $M \times N$ to P will just be called a bilinear mapping.

The reader has already encountered several important examples of bilinear mappings in elementary courses. Here are some of these examples:

Example 12.3 (a) Let V be a vector space over the field of real numbers $\mathbb{R}$. Then any inner product, $\langle\ ,\ \rangle : V \times V \mapsto \mathbb{R}$, on V is clearly an $\mathbb{R}$-bilinear mapping from $V \times V$ to $\mathbb{R}$. For a concrete example, we have the standard inner product, $\langle \alpha, \beta \rangle = \alpha^t\beta$, on $\mathbb{R}^n$.

We note in passing that a complex inner product on a complex vector space V is not a $\mathbb{C}$-bilinear mapping from $V \times V$ to $\mathbb{C}$. (See [2] for definitions.)

(b) If R is a commutative ring, then multiplication $(x, y) \mapsto xy$ is an R-bilinear mapping from $R \times R$ to R. Notice that addition $(x, y) \mapsto x + y$ is not an R-bilinear mapping from $R \times R$ to R.

(c) The function $\psi : R^2 \times R^2 \mapsto R$ given by

$$\psi\left[\begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} u \\ v \end{bmatrix}\right] = \det\begin{bmatrix} x & u \\ y & v \end{bmatrix}$$

is an R-bilinear mapping. ∎

Suppose $\psi : M \times N \mapsto P$ is a bilinear mapping. We take this opportunity to caution the reader about the difference between bilinear mappings and R-module homomorphisms. The product $M \times N$ is an R-module with addition and scalar multiplication given componentwise:

$$(m, n) + (m', n') = (m + m', n + n'), \qquad r(m, n) = (rm, rn)$$

An R-bilinear mapping ψ is rarely an R-module homomorphism from $M \times N$ to P. In fact, suppose the bilinear mapping ψ is also an R-module homomorphism. For any $m \in M$ and $n \in N$, we have $\psi(m, n + n) = \psi(m, n) + \psi(m, n)$ from equation 12.2. Since ψ is also an R-module homomorphism, we have

$$\psi(m, n + n) = \psi((0, n) + (m, n)) = \psi(0, n) + \psi(m, n)$$

Comparing these results, we have $\psi(m, n) = \psi(0, n)$. Again, since ψ is bilinear, $\psi(0, n) = \psi(00, n) = 0\psi(0, n) = 0$. Therefore $\psi(m, n) = 0$. Thus,

the only R-bilinear mapping from $M \times N$ to P which is also an R-module homomorphism is the zero map. In symbols, we have

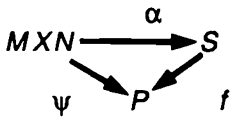
12.4 $\text{Bil}_R(M \times N, P) \cap \text{Hom}_R(M \times N, P) = (0)$

In the theory of bilinear functions, the following universal mapping problem naturally arises.

12.5 (Universal Mapping Problem) Suppose M and N are R-modules. Do there exist an R-module S and an R-bilinear mapping $\alpha : M \times N \mapsto S$ with the following property? Given any R-module P and any R-bilinear mapping $\psi : M \times N \mapsto P$, there exists a unique R-module homomorphism $f : S \mapsto P$ such that $f\alpha = \psi$.

In other words, can we construct an R-module S and an R-bilinear mapping $\alpha : M \times N \mapsto S$ such that any R-bilinear mapping $\psi : M \times N \mapsto P$ factors uniquely through S forming the following commutative diagram?

12.6

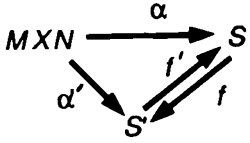


We will see in a moment that the tensor product of M and N is a solution to the universal mapping problem. Before constructing tensor products, we point out that any solution to the universal mapping problem is unique up to isomorphism. To be more precise, we have the following lemma.

Lemma 12.7 Suppose (S, α) and (S', α') are two solutions to the universal mapping problem. Then there exist R-module isomorphisms $f : S \cong S'$ and $f' : S' \cong S$ such that

(a) $f'f = I_S$ and $ff' = I_{S'}$.

(b) The following diagram commutes:

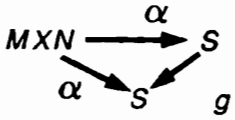


Here I_S and $I_{S'}$ denote the identity maps on S and S', respectively. The diagram in (b) commutes if $f\alpha = \alpha'$ and $f'\alpha' = \alpha$.

Proof. Since (S, α) is a solution to the universal mapping problem and α' is an R-bilinear mapping from $M \times N$ to S', there exists a unique R-module homomorphism $f : S \mapsto S'$ such that $f\alpha = \alpha'$. Similarly, there exists a unique R-module homomorphism $f' : S' \mapsto S$ such that $f'\alpha' = \alpha$. The composite map $f'f : S \mapsto S$ is an R-module homomorphism for which $f'f\alpha = \alpha$.

On the other hand, since (S, α) is a solution to the universal mapping problem, there exists a unique R-module homomorphism $g : S \mapsto S$ such that

12.8



is commutative. Clearly, the R-module homomorphism I_S makes 12.8 commute. Therefore, $g = I_S$. Since $f'f$ also makes 12.8 commute, $f'f = I_S$. A similar proof shows $ff' = I_{S'}$. Thus, the maps f and f' satisfy (a) and (b). ∎

If we construct any solution (S, α) of the universal mapping problem, then Lemma 12.7 implies (S, α) is the only solution to the universal mapping problem (up to isomorphism).

Before constructing a solution to the universal mapping problem, it might be wise to review a few facts about direct sums of R-modules.

Suppose Δ is any nonempty set (finite or infinite). Consider the R-module $U = \bigoplus_{i \in \Delta} R$. Thus, U is the direct sum of Δ copies of R. The module U is just the set of all functions from Δ to R which vanish except possibly at finitely many elements of Δ. In symbols,

$$U = \{h : \Delta \mapsto R \mid h(i) = 0 \text{ except possibly for finitely many } i \in \Delta\}$$

Addition and scalar multiplication on U are defined by the following formulas:

12.9
$$(h_1 + h_2)(i) = h_1(i) + h_2(i) \quad \text{for all } i \in \Delta$$
$$(rh)(i) = r(h(i)) \quad \text{for all } i \in \Delta$$

In equation 12.9, h_1, h_2, and h are functions in U and $r \in R$.

For each index $i \in \Delta$, let $\delta_i : \Delta \mapsto R$ be the function defined by the following formula:

12.10
$$\delta_i(j) = \begin{cases} 0 & \text{if } j \neq i \\ 1 & \text{if } j = i \end{cases} \quad \text{for all } j \in \Delta$$

Clearly, each δ_i is a function in U. If h is any function in U, then h can be written uniquely in the following form:

$$\textbf{12.11} \quad h = \sum_{i \in \Delta} h(i)\delta_i$$

The sum in equation 12.11 makes sense and is, in fact, just some finite linear combination of the vectors in $\{\delta_i \mid i \in \Delta\}$ since $h(i) = 0$ for all but possibly finitely many i in Δ. It follows immediately from equation 12.11 that $\{\delta_i \mid i \in \Delta\}$ is a free R-module basis of U. Thus, $U = \oplus_{i \in \Delta} R$ is a free R-module with free R-module basis $\Gamma = \{\delta_i \mid i \in \Delta\}$. The vectors in Γ are in a one-to-one correspondence with the elements of Δ.

For example, suppose Δ is finite. Say $\Delta = \{1, \ldots, n\}$. We can identify a function $h : \Delta \mapsto R$ with the column vector $h = (h(1), h(2), \ldots, h(n))^t \in R^n$. Then $U = \oplus_{i \in \Delta} R = R^n$, and $\Gamma = \{\delta_i \mid i \in \Delta\} = \varepsilon$, the canonical basis of R^n.

We can use these ideas to construct a solution to the universal mapping problem. Suppose M and N are R-modules. Set $\Delta = M \times N = \{m, n) \mid m \in M, n \in N\}$. Notice that Δ is (usually) a very large set. The elements in Δ are ordered pairs (m, n), $m \in M$, and $n \in N$. Set $U = \oplus_{(m, n) \in \Delta} R$. Thus, U is the direct sum of $M \times N$ copies of R. As we have seen above, U is a free R-module with free R-module basis $\Gamma = \{\delta_{(m,n)} \mid (m, n) \in \Delta = M \times N\}$. Let U_0 denote the R-submodule of U generated by all vectors in U of the following four types:

$$\textbf{12.12:} \quad \begin{aligned} &\delta_{(m+m',n)} - \delta_{(m,n)} - \delta_{(m',n)} \\ &\delta_{(m,n+n')} - \delta_{(m,n)} - \delta_{(m,n')} \\ &\delta_{(rm,n)} - r\delta_{(m,n)} \\ &\delta_{(m,rn)} - r\delta_{(m,n)} \end{aligned}$$

Thus, as m and m', n and n', and r range over M, N, and R, respectively, the elements listed in equation 12.12 form an R-module basis of U_0.

Set $S = U/U_0$. Then S is an R-module. There is a natural map $\alpha : M \times N \mapsto S$ given by $\alpha(m, n) = \delta_{(m,n)} + U_0$. Thus, $\alpha(m, n)$ is just the coset in U/U_0 containing the basis vector $\delta_{(m,n)}$. Since the R-submodule U_0 is generated by the vectors listed in equation 12.12, it is easy to check that α is an R-bilinear mapping from $M \times N$ to S. For example, using the first type of generator in 12.12, we have

$$\begin{aligned} \alpha(m + m', n) &= \delta_{(m+m',n)} + U_0 = (\delta_{(m,n)} + \delta_{(m',n)}) + U_0 \\ &= (\delta_{(m,n)} + U_0) + (\delta_{(m',n)} + U_0) \\ &= \alpha(m, n) + \alpha(m', n) \end{aligned}$$

The other three types of generators in equation 12.12 ensure that α satisfies the remaining three necessary conditions in order that the equations in 12.2 be satisfied. Thus, $\alpha \in \text{Bil}_R(M \times N, S)$.

We claim $(U/U_0, \alpha)$ is a solution to the universal mapping problem. This claim follows from our next lemma.

Lemma 12.13 Let $\psi : M \times N \mapsto P$ be an R-bilinear mapping. Then there exists a unique R-module homomorphism $f : U/U_0 \mapsto P$ such that $f\alpha = \psi$.

Proof. $\Gamma = \{\delta_{(m,n)} \mid (m, n) \in M \times N\}$ is a free R-module basis of $U = \oplus_{(m,n)\in M\times N} R$. Hence, there exists a unique R-module homomorphism $f' : U \mapsto P$ such that $f'(\delta_{(m,n)}) = \psi(m, n)$ for all $(m, n) \in M \times N$. Since U_0 is generated by the vectors given in 12.12, and ψ is a bilinear mapping, f' vanishes on U_0. Hence, f' induces an R-module homomorphism $f : U/U_0 \mapsto P$ given by $f(\delta_{(m,n)} + U_0) = f'(\delta_{(m,n)}) = \psi(m, n)$. In particular, $f\alpha = \psi$.

The fact that f is unique is obvious. ∎

Let us introduce a couple of definitions before summarizing what we have now proved.

Definition 12.14 The R-module U/U_0 constructed above is called the tensor product of M and N (over R) and is written $M \otimes_R N$.

Definition 12.15 The coset $\delta_{(m,n)} + U_0$ in $M \otimes_R N$ is called the tensor product of m and n. We will let $m \otimes_R n$ denote the tensor product of m and n.

When the ring R is clear from the discussion, we will drop the letter R from our notation. Thus, $M \otimes N$ is the tensor product of M and N over R and $m \otimes n$ is the coset $\delta_{(m,n)} + U_0$ in $M \otimes N$. With these changes in notation, the generators of U_0 being zero in U/U_0 imply the following relations in $M \otimes N$:

12.16:
$$(m + m') \otimes n = (m \otimes n) + (m' \otimes n)$$
$$m \otimes (n + n') = (m \otimes n) + (m \otimes n')$$
$$rm \otimes n = m \otimes rn = r(m \otimes n)$$

These relations hold for all $m, m' \in M$, $n, n' \in N$, and $r \in R$.

The bilinear mapping $\alpha : M \times N \mapsto M \otimes N$ is given by $\alpha(m, n) = m \otimes n$ for all $(m, n) \in M \times N$. Lemmas 12.13 and 12.7 can now be put put together in the following theorem.

Theorem 12.17 Let M and N be R-modules. There exist an R-module $M \otimes_R N$ and an R-bilinear mapping $\alpha : M \times N \mapsto M \otimes_R N$ given by $\alpha(m, n) =$

$m \otimes n$ with the following property: If $\psi : M \times N \mapsto P$ is any R-bilinear mapping, then there exists a unique R-module homomorphism $f : M \otimes_R N \mapsto P$ such that $f\alpha = \psi$.

Furthermore, if S' is an R-module and $\alpha' : M \times N \mapsto S'$ an R-bilinear mapping such that (S', α') is a solution to the universal mapping problem, then there exists a unique R-module isomorphism of $j : M \otimes_R N \cong S'$ such that $j\alpha = \alpha'$. ∎

Thus, up to isomorphism, $(M \otimes_R N, \alpha : (m, n) \mapsto m \otimes n)$ is the unique solution to the universal mapping problem. Notice that Theorem 12.17 implies there exists a one-to-one correspondence between $\mathrm{Bil}_R(M \times N, P)$ and $\mathrm{Hom}_R(M \otimes_R N, P)$ given by $\psi \mapsto f$.

Having constructed the tensor product $M \otimes N$ of two modules M and N to solve the universal mapping problem given in 12.5, we can now consider the general problem for n modules $M_1, \ldots, M_n$. A function $\psi : M_1 \times \cdots \times M_n \mapsto P$ is called an R-multilinear mapping if for each $i = 1, \ldots, n$ and for all vectors $m_1, \ldots, m_{i-1}, m_{i+1}, \ldots, m_n$ in $M_1, \ldots, M_{i-1}, M_{i+1}, \ldots, M_n$, respectively, the map

$$\psi(m_1, \ldots, m_{i-1}, *, m_{i+1}, \ldots, m_n) \in \mathrm{Hom}_R(M_i, P)$$

In other words, the function $\psi(x_1, \ldots, x_n) : M_1 \times \cdots \times M_n \mapsto P$ is a multilinear mapping if ψ is a linear function of each variable when the other variables are held fixed. If $n = 2$, a multilinear map is just a bilinear map. We have the analog of the universal mapping problem for multilinear maps:

12.18 Let $M_1, \ldots, M_n$ be R-modules. Do there exist an R-module S and an R-multilinear mapping $\alpha : M_1 \times \cdots \times M_n \mapsto S$ with the following property? For any R-multilinear mapping $\psi : M_1 \times \cdots \times M_n \mapsto P$, there exists a unique R-module homomorphism $f : S \mapsto P$ such that $f\alpha = \psi$.

There is a solution to 12.18 (unique up to isomorphism) which is entirely analogous to the $n = 2$ case. Let U be the following free R-module:

$$U = \bigoplus_{(m_1, \ldots, m_n) \in M_1 \times \cdots \times M_n} R$$

Let U_0 be the R-submodule of U spanned by the analogs of the expressions in equation 12.12, e.g.,

$$\delta_{(m_1, \ldots, m_i + m_i', \ldots, m_n)} - \delta_{(m_1, \ldots, m_i, \ldots, m_n)} - \delta_{(m_1, \ldots, m_i', \ldots, m_n)}$$

Set $S = U/U_0 = M_1 \otimes_R \cdots \otimes_R M_n$. The map $\alpha : M_1 \times \cdots \times M_n \mapsto M_1 \otimes_R \cdots \otimes_R M_n$ is given by

$$\alpha(m_1, \ldots, m_n) = \delta_{(m_1, \ldots, m_n)} + U_0 = m_1 \otimes_R \cdots \otimes_R m_n$$

We then have the analog of Theorem 12.17.

Theorem 12.19 Let $M_1, \ldots, M_n$ be R-modules. There exist an R-module $M_1 \otimes_R \cdots \otimes_R M_n$ and an R-multilinear mapping $\alpha : M_1 \times \cdots \times M_n \mapsto M_1 \otimes_R \cdots \otimes_R M_n$ given by $\alpha(m_1, \ldots, m_n) = m_1 \otimes_R \cdots \otimes_R m_n$ such that the following property is satisfied: If $\psi : M_1 \times \cdots \times M_n \mapsto P$ is any R-multilinear mapping, then there exists a unique R-module homomorphism $f : M_1 \otimes_R \cdots \otimes_R M_n \mapsto P$ such that $f\alpha = \psi$.

Furthermore, if (S', α') is a second solution to the universal mapping problem, then there exists a unique R-module isomorphism $j : M_1 \otimes_R \cdots \otimes_R M_n \mapsto S'$ such that $j\alpha = \alpha'$. ∎

The proof of Theorem 12.19 is the same as the proof of Theorem 12.17. We leave all details to the exercises at the end of this chapter. The R-module $M_1 \otimes_R \cdots \otimes_R M_n$ is called the tensor product of $M_1, \ldots, M_n$. Once again, if R is understood from the context, we will drop R from the notation and write $M_1 \otimes \cdots \otimes M_n$. Notice that $f\alpha = \psi$ in Theorem 12.19 means $f(m_1 \otimes \cdots \otimes m_n) = \psi(m_1, \ldots, m_n)$ for all $(m_1, \ldots, m_n) \in M_1 \times \cdots \times M_n$.

Having introduced the tensor product, let us now discuss some of this construction's more interesting properties.

Theorem 12.20 Let M, N, and P be R-modules. Then the following R-modules are isomorphic:

(a) $M \otimes_R N \cong N \otimes_R M \qquad [m \otimes n \mapsto n \otimes m]$

(b) $(M \otimes_R N) \otimes_R P \cong M \otimes_R (N \otimes_R P) \cong M \otimes_R N \otimes_R P$
$[(m \otimes n) \otimes p \mapsto m \otimes (n \otimes p) \mapsto m \otimes n \otimes p]$

(c) $(M \oplus N) \otimes_R P \cong (M \otimes_R P) \oplus (N \otimes_R P)$
$[(m, n) \otimes p \mapsto (m \otimes p, n \otimes p)]$

(d) $R \otimes_R M \cong M \qquad [r \otimes m \mapsto rm]$

Proof. The maps giving the isomorphisms in (a) through (d) are listed in square brackets at the right of or below each isomorphism. In each case, the isomorphism in question is constructed from the uniqueness of the solution to the universal mapping problem. We illustrate the method by proving (a). We leave the proofs of (b) through (d) to the exercises.

Define a map $\psi : M \times N \mapsto N \otimes_R M$ by $\psi(m, n) = n \otimes m$. Since $n \otimes m = \delta_{(n,m)} + U_0$, ψ is a well-defined function. The relations listed in equation 12.16 (for the R-module $N \otimes_R M$) imply ψ is an R-bilinear mapping. By

Theorem 12.17, there exists a unique R-module homomorphism $f: M \otimes_R N \mapsto N \otimes_R M$ such that $f(m \otimes n) = n \otimes m$ for all $(m, n) \in M \times N$. Similarly, there exists a unique R-module homomorphism $g : N \otimes_R M \mapsto M \otimes_R N$ such that $g(n \otimes m) = m \otimes n$.

Let $\alpha : M \times N \mapsto M \otimes_R N$ be the canonical bilinear mapping given by $\alpha(m, n) = m \otimes n$. Then $gf\alpha = \alpha$ and $I\alpha = \alpha$. Here I denotes the identity map on $M \otimes_R N$. Since $(M \otimes_R N, \alpha)$ is a solution to the universal mapping problem, there is only one R-module homomorphism $p : M \otimes_R N \mapsto M \otimes_R N$ such that $p\alpha = \alpha$, namely $p = I$. Therefore, $gf = I$. Similarly, fg is the identity map on $N \otimes_R M$. In particular, $f(m \otimes n) = n \otimes m$ is an R-module isomorphism of $M \otimes_R N$ onto $N \otimes_R M$. ∎

Let M, M', N, and N' be R-modules. Suppose $f \in \mathrm{Hom}_R(M, M')$ and $g \in \mathrm{Hom}_R(N, N')$. The reader can easily check that $\psi : M \times N \mapsto M' \otimes_R N'$ given by $\psi(m, n) = f(m) \otimes g(n)$ is an R-bilinear mapping. Hence, by Theorem 12.17 there exists a (unique) R-module homomorphism $h : M \otimes_R N \mapsto M' \otimes_R N'$ such that $h(m \otimes n) = f(m) \otimes g(n)$ for all $(m, n) \in M \times N$. The R-module homomorphism h is called the tensor product of f and g. The standard notation for the tensor product of f and g is $f \otimes g$. Thus, $f \otimes g \in \mathrm{Hom}_R(M \otimes_R N, M' \otimes_R N')$, and $(f \otimes g)(m \otimes n) = f(m) \otimes g(n)$ for all $m \in M$ and $n \in N$.

Composites of homomorphisms work nicely with respect to the tensor product. Suppose

$$M \overset{f}{\mapsto} M' \overset{f'}{\mapsto} M'' \quad \text{and} \quad N \overset{g}{\mapsto} N' \overset{g'}{\mapsto} N''$$

are R-modules and R-module homomorphisms. Then the following identity is true.

12.21 $(f'f) \otimes (g'g) = (f' \otimes g')(f \otimes g)$

The R-module homomorphisms on each side of equation 12.21 are maps from $M \otimes_R N$ to $M'' \otimes_R N''$. We have seen in the construction of the tensor product that $M \otimes_R N$ is generated as an R-module by all elements of the form $m \otimes n$ with $(m, n) \in M \times N$. In particular, the two R-module homomorphisms given in equation 12.21 are equal if they have the same value on $m \otimes n$ for all $(m, n) \in M \times N$. Since

$$\begin{aligned}[(f'f) \otimes (g'g)](m \otimes n) &= (f'f)(m) \otimes (g'g)(n) \\ &= f'(f(m)) \otimes g'(g(n)) = (f' \otimes g')(f(m) \otimes g(n)) \\ &= [(f' \otimes g')(f \otimes g)](m \otimes n)\end{aligned}$$

the equality in 12.21 is established.

One of the most important properties of the tensor product is the fact that $M \otimes_R (*)$ and $(*) \otimes_R M$ are right exact functors. This means if

$$N' \xrightarrow{g} N \xrightarrow{f} N'' \to (0)$$

is a right exact sequence of R-modules, then

$$\textbf{12.22} \quad M \otimes_R N' \xrightarrow{1\otimes g} M \otimes_R N \xrightarrow{1\otimes f} M \otimes_R N'' \to (0)$$

$$N' \otimes_R M \xrightarrow{g\otimes 1} N \otimes_R M \xrightarrow{f\otimes 1} N'' \otimes_R M \to (0)$$

are right exact sequences. In the sequences in 12.22, 1 denotes the identity map from M to M. By Theorem 12.20a, it suffices to prove one of these sequences is right exact. Consequently, we prove the following result.

Theorem 12.23 Suppose

$$N' \xrightarrow{g} N \xrightarrow{f} N'' \to (0)$$

is a right exact sequence of R-modules. Let M be any R-module. Then

$$M \otimes_R N' \xrightarrow{1\otimes g} M \otimes_R N \xrightarrow{1\otimes f} M \otimes_R N'' \to (0)$$

is a right exact sequence of R-modules.

Proof. We have seen that $G = \{m \otimes n'' \mid m \in M, n'' \in N''\}$ is a set of R-module generators of $M \otimes_R N''$. Fix an element $m \otimes n'' \in G$. Since f is a surjective, R-module homomorphism, there exists an $n \in N$ such that $f(n) = n''$. Therefore, $(1 \otimes f)(m \otimes n) = m \otimes n''$. Thus, $G \subseteq \text{Im}(1 \otimes f)$. This immediately implies $(1 \otimes f)(M \otimes_R N) = M \otimes_R N''$; that is, $1 \otimes f$ is surjective.

By equation 12.21, $(1 \otimes f)(1 \otimes g) = 1 \otimes fg = 1 \otimes 0 = 0$. Therefore, $\text{Im}(1 \otimes g) \subseteq \text{Ker}(1 \otimes f)$. Thus, to complete the proof of the theorem, we must argue $\text{Ker}(1 \otimes f) \subseteq \text{Im}(1 \otimes g)$. Since $\text{Im}(1 \otimes g) \subseteq \text{Ker}(1 \otimes f)$, there is a natural R-module surjection $M \otimes N/\text{Im}(1 \otimes g) \mapsto M \otimes N/\text{Ker}(1 \otimes f)$. Since $1 \otimes f$ is surjective, the first isomorphism theorem implies $M \otimes N/\text{Ker}(1 \otimes f) \cong M \otimes N''$. Putting these two maps together, we have an R-module homomorphism $h : M \otimes N/\text{Im}(1 \otimes g) \mapsto M \otimes N''$. The map h is given by $h(m \otimes n + \text{Im}(1 \otimes g)) = m \otimes f(n)$ for all $m \in M$ and $n \in N$. Since the two maps making up h are surjective, h is surjective. It is clear from the definition of h that h is an isomorphism if and only if $\text{Im}(1 \otimes g) = \text{Ker}(1 \otimes f)$. Thus, it suffices to argue h is an isomorphism.

Suppose $m \in M$ and $n'' \in N''$. Since f is surjective, there exists an $n \in N$ such that $f(n) = n''$. Then $h(m \otimes n + \text{Im}(1 \otimes g)) = m \otimes f(n) = m \otimes n''$. Suppose $n_1 \in N$ is another element of N such that $f(n_1) = n''$. Then $f(n - n_1)$

$= f(n) - f(n_1) = n'' - n'' = 0$. Therefore, $n - n_1 \in \text{Ker}(f) = \text{Im}(g)$. In particular, $m \otimes (n - n_1) \in \text{Im}(1 \otimes g)$. Thus, the cosets $m \otimes n + \text{Im}(1 \otimes g)$ and $m \otimes n_1 + \text{Im}(1 \otimes g)$ are equal in $M \otimes_R N/\text{Im}(1 \otimes g)$. Hence, there is a well-defined function $\psi : M \times N'' \mapsto M \otimes_R N/\text{Im}(1 \otimes g)$ given by $\psi(m, n'') = m \otimes n + \text{Im}(1 \otimes g)$. Here n is any element in N such that $f(n) = n''$. It is easy to check that ψ is an R-bilinear mapping from $M \times N''$ to $M \otimes_R N/\text{Im}(1 \otimes g)$. Theorem 12.17 implies there exists a unique R-module homomorphism $p : M \otimes_R N'' \mapsto M \otimes_R N/\text{Im}(1 \otimes g)$ such that $p(m \otimes n'') = \psi(m, n'') = m \otimes n + \text{Im}(1 \otimes g)$.

Now

$$\begin{aligned} hp(m \otimes n'') &= h\psi(m, n'') = h(m \otimes n + \text{Im}(1 \otimes g)) \\ &= m \otimes f(n) = m \otimes n'' \end{aligned}$$

for any $n \in N$ such that $f(n) = n''$. Also,

$$\begin{aligned} ph(m \otimes n + \text{Im}(1 \otimes g)) &= p(m \otimes f(n)) \\ &= \psi(m, f(n)) = m \otimes n + \text{Im}(1 \otimes g) \end{aligned}$$

Thus, the R-module homomorphisms h and p are inverses of each other. In particular, h is an isomorphism and $\text{Im}(1 \otimes g) = \text{Ker}(1 \otimes f)$. ∎

We note that the functor $M \otimes_R (*)$ does not in general preserve injective maps. Thus, $M \otimes_R (*)$ does not in general preserve short exact sequences. Consider the following example.

Example 12.24 Consider the following short exact sequence of $\mathbb{Z}$-modules:

$$(0) \mapsto \mathbb{Z} \overset{g}{\mapsto} \mathbb{Z} \overset{f}{\mapsto} \mathbb{Z}/2\mathbb{Z} \mapsto (0)$$

Here g is the $\mathbb{Z}$-module homomorphism given by $g(x) = 2x$ and f is the natural map given by $f(x) = x + 2\mathbb{Z}$. Suppose we tensor this sequence with the $\mathbb{Z}$-module $\mathbb{Z}/2\mathbb{Z}$. By Theorem 12.23, we get the following right exact sequence:

$$\textbf{12.25} \quad \mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z} \overset{1 \otimes g}{\mapsto} \mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z} \overset{1 \otimes f}{\mapsto} \mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/2\mathbb{Z} \mapsto (0)$$

The sequence in 12.25 is not a short exact sequence of $\mathbb{Z}$-modules because the $\mathbb{Z}$-module homomorphism $1 \otimes g$ is not injective. To see this, first note that Theorem 12.20d implies $\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z} \neq (0)$. If $x = \bar{0}$ or $\bar{1}$ in $\mathbb{Z}/2\mathbb{Z}$ and $n \in \mathbb{Z}$, then

$$(1 \otimes g)(x \otimes n) = x \otimes g(n) = x \otimes 2n = 2x \otimes n = 0 \otimes n = 0$$

Thus, the map $1 \otimes g$ is identically zero. ∎

The tensor product is often used to extend scalars from one ring to another. Suppose M is an R-module and $f: R \mapsto T$ a ring homomorphism (of commutative rings). Then T is an R-module with scalar multiplication given by $rt = f(r)t$ for all $r \in R$ and $t \in T$. In particular, we can form the tensor product $M \otimes_R T$ of the R-modules M and T. It is easy to see that the R-module $M \otimes_R T$ is a T-module with scalar multiplication defined by $t(m \otimes t') = m \otimes tt'$.

If $\Gamma = \{m_1, \ldots, m_n\}$ is an R-module basis of M, then clearly $\{m_1 \otimes 1, \ldots, m_n \otimes 1\}$ is a T-module basis of $M \otimes_R T$. Thus, if M is a finitely generated R-module, $M \otimes_R T$ is a finitely generated T-module.

The T-module $M \otimes_R T$ is said to be obtained from M by extending scalars (from R to T). Thus, the property of being finitely generated is preserved under extension of scalars. On the other hand, extending scalars usually does not preserve short exact sequences. If

$$(0) \mapsto N' \overset{g}{\mapsto} N \overset{f}{\mapsto} N'' \mapsto (0)$$

is a short exact sequence of R-modules, then

$$N' \otimes_R T \overset{g\otimes 1}{\mapsto} N \otimes_R T \overset{f\otimes 1}{\mapsto} N'' \otimes_R T \mapsto (0)$$

is a right exact sequence of T-modules by Theorem 12.23. The map $g \otimes 1$ need not be injective, as Example 12.24 shows.

EXERCISES

1. Let V be a vector space over the complex numbers $\mathbb{C}$. Why is a complex inner product on V not a $\mathbb{C}$-bilinear mapping?
2. In the construction of the tensor product, complete the argument that $\alpha(m, n) = \delta_{(m,n)} + U_0$ is an R-bilinear mapping of $M \times N$ into U/U_0.
3. In the proof of Lemma 12.13, show that $f(\delta_{(m,n)} + U_0) = \psi(m, n)$ is the only R-module homomorphism from U/U_0 to P for which $f\alpha = \psi$.
4. Prove Theorem 12.19.
5. Let $\mathrm{Mul}_R(M_1 \times \cdots \times M_n, P)$ denote the set of all R-multilinear mappings of $M_1 \times \cdots \times M_n$ into P. Prove the following assertions:
   (a) $\mathrm{Mul}_R(M_1 \times \cdots \times M_n, P) \cap \mathrm{Hom}_R(M_1 \times \cdots \times M_n, P) = (0)$.
   (b) There is a 1-1 correspondence between the two sets

   $\mathrm{Mul}_R(M_1 \times \cdots \times M_n, P)$ and $\mathrm{Hom}_R(M_1 \times \cdots \times M_n, P)$.
6. In Exercise 5, if $\psi \in \mathrm{Mul}_R(M_1 \times \cdots \times M_n, P)$, is $\mathrm{Im}(\psi)$ an R-submodule of P?
7. Prove (b), (c), and (d) in Theorem 12.20.
8. Generalize Theorem 12.20c as follows: Suppose $\{M_i \mid i \in \Delta\}$ is a collection of R-modules. Show $(\oplus_{i\in\Delta} M_i) \otimes_R P \cong \oplus_{i\in\Delta}(M_i \otimes_R P)$.

9. Suppose $f: R \mapsto T$ is a homomorphism of commutative rings. Let M be an R-module. Show that the scalar multiplication $t(m \otimes t') = m \otimes tt'$ is well defined and makes $M \otimes_R T$ a T-module.
10. With the same notation as in Exercise 9, suppose $M_1, \ldots, M_n$ are R-modules. Show the T-module $(M_1 \otimes_R T) \otimes_T \cdots \otimes_T (M_n \otimes_R T)$ is isomorphic to the T-module $(M_1 \otimes_R \cdots \otimes_R M_n) \otimes_R T$.
11. If M and N are free R-modules of ranks m and n, respectively, show $M \otimes_R N$ is a free R-module of rank mn.
12. If $\mathfrak{A}$ is an ideal in R and M is an R-module, show $M \otimes_R R/\mathfrak{A} \cong M/\mathfrak{A}M$.
13. Let $\mathfrak{A}$ and $\mathfrak{B}$ be ideals in R. Prove the following assertions:
    (a) The R-module $(R/\mathfrak{A}) \otimes_R (R/\mathfrak{B})$ is a commutative ring with multiplication defined as follows: $(x \otimes y)(x' \otimes y') = xx' \otimes yy'$.
    (b) $(R/\mathfrak{A}) \otimes_R (R/\mathfrak{B}) \cong R/(\mathfrak{A} + \mathfrak{B})$ (as rings).
    (c) $\mathbb{Z}/n\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/m\mathbb{Z} \cong \mathbb{Z}/d\mathbb{Z}$ where d is a greatest common divisor of m and n.
14. Suppose R is a local ring $R = (R, m, k)$. Let M and N be finitely generated R-modules. Show $M \otimes_R N = (0)$ if and only if $M = (0)$ or $N = (0)$.
15. Let M and N be R-modules. Suppose $\sum_{i=1}^{n} x_i \otimes y_i = 0$ in $M \otimes_R N$. Show there exist finitely generated submodules $M_0 \subseteq M$ and $N_0 \subseteq N$ such that $\sum_{i=1}^{n} x_i \otimes y_i = 0$ in $M_0 \otimes_R N_0$.
16. Suppose M is a direct summand of a free R-module F. Show $M \otimes_R (*)$ preserves short exact sequences. Thus, if

$$(0) \mapsto N' \overset{g}{\mapsto} N \overset{f}{\mapsto} N'' \mapsto (0)$$

is a short exact sequence of R-modules, then

$$(0) \mapsto M \otimes_R N' \overset{1\otimes g}{\mapsto} M \otimes_R N \overset{1\otimes f}{\mapsto} M \otimes_R N'' \mapsto (0)$$

is also a short exact sequence of R-modules.
17. Suppose

$$(0) \mapsto N' \overset{g}{\mapsto} N \overset{f}{\mapsto} N'' \mapsto (0)$$

$$(0) \mapsto M' \overset{h}{\mapsto} M \overset{k}{\mapsto} M'' \mapsto (0)$$

are short exact sequences of R-modules. Show the complex of R-modules

$$(0) \mapsto N' \otimes_R M' \overset{g\otimes h}{\mapsto} N \otimes_R M \overset{f\otimes k}{\mapsto} N'' \otimes_R M'' \mapsto (0)$$

is not necessarily exact even when R is a field.
18. Let F be a field. Let V and W be finite-dimensional vector spaces over F, and let $T \in \mathrm{Hom}_F(V, V)$ and $S \in \mathrm{Hom}_F(W, W)$. If A and B are matrix

representations of T and S, respectively, find a matrix representation of $T \otimes S \in \mathrm{Hom}_F(V \otimes_F W, V \otimes_F W)$ in terms of A and B. (Hint: Consider the Kronecker product of A and B. See [2; p. 67].)

(d) Let M, N, and P be R-modules. Show

$$\mathrm{Hom}_R(M \otimes_R N, P) \cong \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, P))$$

19. Suppose P is a direct summand of a finitely generated, free R-module F. Show

$$\mathrm{Hom}_R(P, M) \otimes_R N \cong \mathrm{Hom}_R(P, M \otimes_R N).$$

20. Suppose F is a field. Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be indeterminates over F. Let $\mathfrak{A}$ be an ideal in $F[X_1, \ldots, X_n]$, and $\mathfrak{B}$ an ideal in $F[Y_1, \ldots, Y_m]$. Show that the two rings $F[X_1, \ldots, X_n]/\mathfrak{A} \otimes_F F[Y_1, \ldots, Y_m]/\mathfrak{B}$ and $F[X_1, \ldots, X_n, Y_1, \ldots, Y_m]/\mathfrak{A} + \mathfrak{B}$ are isomorphic as F-algebras. (Before proving this theorem, give some thought to what the symbols $\mathfrak{A} + \mathfrak{B}$ mean as an ideal in $F[X_1, \ldots, X_n, Y_1, \ldots, Y_m]$.)

13
Fitting Ideals

In this chapter, we will assume R is a Noetherian ring and M a finitely generated R-module. Suppose $\Gamma = \{m_1, \ldots, m_n\}$ is an R-module basis of M. If $M = (0)$, then we take $n = 1$ and $m_1 = 0$. Every element $m \in M$ can be written as an R-linear combination of the elements of Γ : $m = r_1m_1 + \cdots + r_nm_n$. Here $r_1, \ldots, r_n \in R$. Since Γ is not necessarily a free R-module basis of M (M may not even be a free R-module), there may be many different R-linear combinations of $m_1, \ldots, m_n$ which give m. We construct a relations matrix for M as in Chapter 10.

Let $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_n\}$ be a free R-module basis of R^n. Let f denote the R-module homomorphism in $\mathrm{Hom}_R(R^n,M)$ determined by $\boldsymbol{\lambda}$ and Γ. Thus, $f(\sum_{i=1}^n r_i\lambda_i) = \sum_{i=1}^n r_im_i$. Let $K = \mathrm{Ker}(f)$. Then

$$\textbf{13.1} \qquad (0) \mapsto K \overset{\iota}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$$

is a short exact sequence of R-modules. The map ι in 13.1 is the inclusion map of K into R^n. Since R is a Noetherian ring, K is a finitely generated R-module. This follows from Corollary 10.8 and Lemma 10.2. Suppose $\Delta = \{\delta_1, \ldots, \delta_m\}$ is an R-module basis of K. Notice that $K = (0)$ if and only if Γ is a free R-module basis of M. In this case, we take $m = 1$ and $\delta_1 = 0$. If we now let $\boldsymbol{\mu}$

$= \{\mu_1, \ldots, \mu_m\}$ be any free R-module basis of R^m and define $g : R^m \mapsto K$ by $g(\sum_{j=1}^m r_j\mu_j) = \sum_{j=1}^m r_j\delta_j$, then $\mathrm{Im}(g) = K$, and

$$\textbf{13.1}' \quad R^m \overset{g}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$$

is a finite presentation of M.

Obviously, the sequences in 13.1 and 13.1′ are closely related. We can use either sequence to define a relations matrix for M. The three bases Γ, $\boldsymbol{\lambda}$, and Δ determine a relations matrix $C(\Gamma,\boldsymbol{\lambda},\Delta) \in M_{m\times n}(R)$ which is defined as follows:

$$\textbf{13.2} \qquad C(\Gamma,\boldsymbol{\lambda},\Delta) = (c_{ij}) \text{ where } \delta_i = \sum_{j=1}^n c_{ij}\lambda_j \text{ for all } j = 1, \ldots, m.$$

Notice that we are regarding Γ, $\boldsymbol{\lambda}$, and Δ as ordered sets here. Any reordering of the elements in any one of these sets results in a new relations matrix. If the elements in Γ ($\boldsymbol{\lambda}$ or Δ) are reordered, we will treat the new set as a different (ordered) basis of M (R^n or K) and write Γ' ($\boldsymbol{\lambda}'$ or Δ') for this set.

Having chosen (ordered) bases Γ, $\boldsymbol{\lambda}$, and Δ for M, R^n, and K, respectively, we then have the relations matrix $C(\Gamma,\boldsymbol{\lambda},\Delta) \in M_{m\times n}(R)$. We can then consider the sequence of ideals $\{I_t(C(\Gamma,\boldsymbol{\lambda},\Delta)) \mid t \in \mathbb{Z}\}$ in R. The reader will recall from Chapter 4 that for $1 \le t \le \min\{m,n\}$, $I_t(C(\Gamma,\boldsymbol{\lambda},\Delta))$ is the ideal in R generated by all $t\times t$ minors of $C(\Gamma,\boldsymbol{\lambda},\Delta)$. If $t \le 0$, then $I_t(C(\Gamma,\boldsymbol{\lambda},\Delta)) = R$, and if $t > \min\{m,n\}$, then $I_t(C(\Gamma,\boldsymbol{\lambda},\Delta)) = (0)$. Thus, we have the following descending sequence of ideals in R:

$$\textbf{13.3} \quad R = I_0(C(\Gamma,\boldsymbol{\lambda},\Delta)) \supseteq I_1(C(\Gamma,\boldsymbol{\lambda},\Delta)) \supseteq I_2(C(\Gamma,\boldsymbol{\lambda},\Delta)) \supseteq \cdots$$

As our notation indicates, the sequence of ideals constructed in 13.3 a priori depends on the choice of ordered bases Γ, $\boldsymbol{\lambda}$, and Δ. In our first lemma in this section, we show the ideals in 13.3 in fact do not depend on any specific choice of Δ.

Lemma 13.4 Suppose Δ and Δ' are two sets of generators of K. Then for every $t \in \mathbb{Z}$, $I_t(C(\Gamma,\boldsymbol{\lambda},\Delta)) = I_t(C(\Gamma,\boldsymbol{\lambda},\Delta'))$.

Proof. Suppose $\Delta = \{\delta_1, \ldots, \delta_m\}$ and $\Delta' = \{\delta_1', \ldots, \delta_p'\}$. Set $C = C(\Gamma,\boldsymbol{\lambda},\Delta) = (c_{ij}) \in M_{m\times n}(R)$. Then $\delta_i = \sum_{j=1}^n c_{ij}\lambda_j$ for all $i = 1, \ldots, m$. Set $D = C(\Gamma,\boldsymbol{\lambda},\Delta') = (d_{kj}) \in M_{p\times n}(R)$. Then $\delta_k' = \sum_{j=1}^n d_{kj}\lambda_j$ for all $k = 1, \ldots, p$. Since Δ and Δ' are both R-module bases of K, we have

$$\textbf{13.5} \quad \delta_i = \sum_{k=1}^p x_{ik}\delta_k' \quad \text{for } i = 1, \ldots, m$$

and

$$\delta_k' = \sum_{i=1}^m y_{ki}\delta_i \quad \text{for } k = 1, \ldots, p$$

The coefficients x_{ik} and y_{ki} in equation 13.5 are elements from R. Set $X = (x_{ik}) \in M_{m\times p}(R)$ and $Y = (y_{ki}) \in M_{p\times m}(R)$. Then for each $i = 1, \ldots, m$, we have

$$\sum_{j=1}^{n} c_{ij}\lambda_j = \delta_i = \sum_{k=1}^{p} x_{ik}\,\delta_k' = \sum_{k=1}^{p} x_{ik}\left(\sum_{j=1}^{n} d_{kj}\lambda_j\right) = \sum_{j=1}^{n}\left(\sum_{k=1}^{p} x_{ik}d_{kj}\right)\lambda_j$$

Since $\boldsymbol{\lambda}$ is a free R-module basis of R^n, we conclude that $c_{ij} = \sum_{k=1}^{p} x_{ik}d_{kj}$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$. A similar computation can be done beginning with δ_k'. Hence, we have

13.6 $C = XD$ and $D = YC$.

It now follows from Lemma 4.5 that $I_t(C) = I_t(XD) \subseteq I_t(D)$ and $I_t(D) = I_t(YC) \subseteq I_t(C)$. Therefore, $I_t(C(\Gamma,\boldsymbol{\lambda},\Delta)) = \mathrm{I}_t(\mathrm{C}(\Gamma,\boldsymbol{\lambda},\Delta'))$ for all $t \in \mathbb{Z}$. ∎

Since the ideals in 13.3 do not depend on any specific choice of basis Δ of K, we will drop the symbol Δ from our notation. Thus, the ideals in 13.3 will be written as follows:

13.7 $R = I_0(C(\Gamma,\boldsymbol{\lambda})) \supseteq \mathrm{I}_1(\mathrm{C}(\Gamma,\boldsymbol{\lambda})) \supseteq \mathrm{I}_2(\mathrm{C}(\Gamma,\boldsymbol{\lambda})) \supseteq \cdots$

We next argue that the ideals in 13.7 do not depend on our choice of the free R-module basis $\boldsymbol{\lambda}$. By this we mean the following: Suppose $\boldsymbol{\lambda}' = \{\lambda_1', \ldots, \lambda_n'\}$ is another free R-module basis of R^n. The bases Γ and $\boldsymbol{\lambda}'$ define an R-module homomorphism $f' : R^n \mapsto M$ given by $f'\left(\sum_{i=1}^{n} r_i\lambda_i'\right) = \sum_{i=1}^{n} r_i m_i$. If $K' = \mathrm{Ker}(f')$, then we get another short exact sequence.

13.8 $(0) \mapsto K' \overset{\iota}{\mapsto} R^n \overset{f'}{\mapsto} M \mapsto (0)$

By choosing a basis Δ' of K', we get another relations matrix $C(\Gamma,\boldsymbol{\lambda}',\Delta')$. We then have two sets of ideals $\{I_t(C(\Gamma,\boldsymbol{\lambda}))\}$ and $\{I_t(C(\Gamma,\boldsymbol{\lambda}'))\}$. Since these ideals do not depend on our choice of Δ and Δ', we drop this part of the notation. We claim these two sets of ideals are exactly the same. This follows from the next lemma.

Lemma 13.9 $I_t(C(\Gamma,\boldsymbol{\lambda})) = \mathrm{I}_t(\mathrm{C}(\Gamma,\boldsymbol{\lambda}'))$ for all $t \in \mathbb{Z}$.

Proof. By Lemma 13.4, we can compute $I_t(C(\Gamma,\boldsymbol{\lambda}'))$ by choosing any basis Δ' of K'. We do this in a special way. Since $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ are both free R-module bases

of R^n, there exists an R-isomorphism $\theta : R^n \mapsto R^n$ given by $\theta(\sum_{i=1}^n r_i\lambda_i) = \sum_{i=1}^n r_i\lambda_i'$. We than have the following diagram.

13.10
$$\begin{array}{ccccccccc} (0) & \mapsto & K & \overset{\iota}{\mapsto} & R^n & \overset{f}{\mapsto} & M & \mapsto & (0) \\ & & & & \theta\downarrow & & \| & & \\ (0) & \mapsto & K' & \overset{\iota}{\mapsto} & R^n & \overset{f'}{\mapsto} & M & \mapsto & (0) \end{array}$$

The square in the right portion of 13.10 commutes; that is, $f'\theta = f$.

We claim that θ induces an R-module isomorphism of K onto K'. To see this, let $\delta \in K$. Then $f'\theta(\delta) = f(\delta) = 0$. Therefore, $\theta(\delta) \in \mathrm{Ker}(f') = K'$. Consequently, $\theta(K) \subseteq K'$. Conversely, suppose $\delta' \in K'$. Then $f'(\delta') = 0$. Since θ is surjective, $\delta' = \theta(\delta)$ for some $\delta \in R^n$. Then $f(\delta) = f'\theta(\delta) = f'(\delta') = 0$. Therefore, $\delta \in \mathrm{Ker}(f) = K$. In particular, $\theta(K) = K'$, and our claim is proved.

Now let $\Delta = \{\delta_1, \ldots, \delta_m\}$ be an R-module basis of K. Let $\delta_i = \sum_{j=1}^n c_{ij}\lambda_j$ for all $i = 1, \ldots, m$. Then $I_t(C(\Gamma,\boldsymbol{\lambda})) = I_t((c_{ij}))$. Since $\theta(K) = K'$, $\Delta' = \{\theta(\delta_1), \ldots, \theta(\delta_m)\}$ is an R-module basis of K'. Also, $\theta(\delta_i) = \sum_{j=1}^n c_{ij}\,\theta(\lambda_j) = \sum_{j=1}^n c_{ij}\lambda_j'$ for all $i = 1, \ldots, m$. Therefore, $I_t(C(\Gamma,\boldsymbol{\lambda}')) = I_t((c_{ij}))$. In particular, $I_t(C(\Gamma,\boldsymbol{\lambda})) = I_t(C(\Gamma,\boldsymbol{\lambda}'))$. This completes the proof of Lemma 13.9. ∎

Since the ideals in 13.3 do not depend on any specific choice of $\boldsymbol{\lambda}$ or Δ, we will drop both symbols from our notation. Hence, we can write the descending chain of ideals in 13.3 as follows:

13.11 $$R = I_0(C(\Gamma)) \supseteq I_1(C(\Gamma)) \supseteq I_2(C(\Gamma)) \supseteq \cdots$$

The absence of notation here means the following: To construct the ideals in 13.11, choose an (ordered) basis Γ of M. Then choose any free R-module basis $\boldsymbol{\lambda}$ of R^n and set up the short exact sequence in 13.1. Choose any basis Δ of $K = \mathrm{Ker}(f)$ and form $C(\Gamma,\boldsymbol{\lambda},\Delta)$. The ideals $\{I_t(C(\Gamma,\boldsymbol{\lambda},\Delta)) \mid t \in \mathbb{Z}\}$ depend only on our first choice Γ.

We have to be careful about the role Γ plays in determining the ideals $\{I_t(C(\Gamma)) \mid t \in \mathbb{Z}\}$. As we will see in Example 13.15, the ideal $I_t(C(\Gamma))$ is dependent on the specific choice of Γ. However, there is one important change in Γ which leaves the ideals in 13.11 the same.

Suppose $\Gamma = \{m_1, \ldots, m_n\}$ is a fixed set of generators of M. Let $X = (x_{ij}) \in \mathrm{Gl}(n,R)$. Set $X\Gamma = \{m_1', \ldots, m_n'\}$ where $m_i' = \sum_{j=1}^n x_{ij}m_j$ for $i = 1, \ldots, m$. Since X is invertible, $X\Gamma$ is another set of generators of M.

Lemma 13.12 $I_t(C(\Gamma)) = I_t(C(X\Gamma))$ for all $t \in \mathbb{Z}$.

Proof. Suppose $\varepsilon = \{\varepsilon_1, \ldots, \varepsilon_n\}$ is the canonical basis of R^n. Then we have the following short exact sequence.

$$\textbf{13.13}\quad (0) \mapsto K \overset{\iota}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$$

In 13.13, f is the R-module homomorphism given by $f(\sum_{i=1}^n r_i\varepsilon_i) = \sum_{i=1}^n r_i m_i$ and $K = \operatorname{Ker}(f)$, As usual, ι is the inclusion map. Let $\Delta = \{\delta_1, \ldots, \delta_m\}$ be an R-module basis of K, and set $\delta_i = \sum_{j=1}^n c_{ij}\varepsilon_j$ for $i = 1, \ldots, m$. The last two lemmas imply $I_t(C(\Gamma)) = I_t((c_{ij}))$ for all $t \in \mathbb{Z}$.

In order to compute $I_t(C(X\Gamma))$, we can choose any free R-module basis of R^n and the corresponding map $f' : R^n \mapsto M$. Let $\varepsilon_k' = \sum_{j=1}^n x_{kj}\varepsilon_j$ for $k = 1, \ldots, n$. Since X is an invertible matrix, $\varepsilon' = \{\varepsilon_1', \ldots, \varepsilon_n'\}$ is a free R-module basis of R^n. This follows from Corollary 5.16. Define $f' : R^n \mapsto M$ by $f'(\sum_{i=1}^n r_i\varepsilon_i') = \sum_{i=1}^n r_i m_i'$. Set $K' = \operatorname{Ker}(f')$. Then

$$\textbf{13.14}\quad (0) \mapsto K' \overset{\iota}{\mapsto} R^n \overset{f'}{\mapsto} M \mapsto (0)$$

is a short exact sequence of R-modules. Notice that

$$f(\varepsilon_k') = f\left(\sum_{j=1}^n x_{kj}\varepsilon_j\right) = \sum_{j=1}^n x_{kj}f(\varepsilon_j) = \sum_{j=1}^n x_{kj}m_j = m_k' = f'(\varepsilon_k')$$

for each $k = 1, \ldots, n$. Since ε' is a basis of R^n, it follows that $f = f'$. In particular, $K = K'$ and Δ is a basis of K'.

Suppose $\delta_i = \sum_{j=1}^n d_{ij}\varepsilon_j'$ for each $i = 1, \ldots, m$. Then $C(X\Gamma) = (d_{ij})$. For each $i = 1, \ldots, m$, we have

$$\sum_{j=1}^n c_{ij}\varepsilon_j = \delta_i = \sum_{k=1}^n d_{ik}\varepsilon_k' = \sum_{k=1}^n d_{ik}\left(\sum_{j=1}^n x_{kj}\varepsilon_j\right) = \sum_{j=1}^n\left(\sum_{k=1}^n d_{ij}x_{kj}\right)\varepsilon_j.$$

Therefore, $c_{ij} = \sum_{k=1}^n d_{ik}x_{kj}$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$. These equations imply $C(\Gamma) = C(X\Gamma)X$. It now follows from Corollary 4.8 that $I_t(C(\Gamma)) = I_t(C(X\Gamma))$ for all $t \in \mathbb{Z}$. ∎

One important application of Lemma 13.12 occurs when X is a permutation matrix. In this case, $X\Gamma$ is just some reordering of the elements of Γ. Lemma 13.12 implies the ideals in 13.11 do not depend on the ordering of the elements in Γ.

We have mentioned that the ideals $\{I_t(C(\Gamma)) \mid t \in \mathbb{Z}\}$ do in general depend on our choice of Γ. Consider the following example.

Example 13.15 Let $R = \mathbb{Z}$, and $M = \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$. Clearly, M is a finitely generated R-module. One basis for M which is readily apparent is $\Gamma = \{m_1 = (\bar{1},0), m_2 = (0,\bar{1})\}$. Then

$$\textbf{13.16}\quad (0) \mapsto K \overset{\iota}{\mapsto} R^2 \overset{f}{\mapsto} M \mapsto (0)$$

is a short exact sequence. Here f is defined by $f(r_1\varepsilon_1 + r_2\varepsilon_2) = r_1m_1 + r_2m_2 = (\bar{r}_1,\bar{r}_2)$ where $\bar{r}$ denotes the image of r in $\mathbb{Z}/2\mathbb{Z}$. $K = \mathrm{Ker}(f)$. It is easy to check that $\Delta = \{2\varepsilon_1, 2\varepsilon_2\}$ is an R-module basis of K. Therefore,

$$C(\Gamma) = C(\Gamma,\boldsymbol{\varepsilon},\Delta) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \in M_{2\times 2}(\mathbb{Z})$$

In particular, $I_1(C(\Gamma)) = 2\mathbb{Z}$.

Suppose $\Gamma' = \{m_1 = (\bar{1},0), m_2 = (0,\bar{1}), m_3 = (\bar{1},\bar{1})\}$. Clearly, Γ' is also an R-module basis of M. The corresponding short exact sequence for Γ' is

$$\textbf{13.17}\quad (0) \mapsto K' \overset{\iota}{\mapsto} R^3 \overset{g}{\mapsto} M \mapsto (0)$$

The map $g : R^3 \mapsto M$ is given by

$$g(r_1\varepsilon_1 + r_2\varepsilon_2 + r_3\varepsilon_3) = r_1m_1 + r_2m_2 + r_3m_3$$

and $K' = \mathrm{Ker}(g)$. A simple calculation shows $\Delta' = \{2\varepsilon_1, 2\varepsilon_2, \varepsilon_1 + \varepsilon_2 + \varepsilon_3\}$ is an R-module basis of K'. Thus,

$$C(\Gamma') = C(\Gamma',\boldsymbol{\varepsilon},\Delta') = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix} \in M_{3\times 3}(R)$$

Therefore, $I_1(C(\Gamma')) = R$. In particular, $I_1(C(\Gamma)) \neq I_1(C(\Gamma'))$. ∎

We can "fix" the situation in Example 13.15 by introducing the following definition.

Definition 13.18 Let $\Gamma = \{m_1, \ldots, m_n\}$ be a set of R-module generators of M. For each integer k, set

$$\mathfrak{F}_k(\Gamma)(M) = \begin{cases} (0) & \text{if } k < 0 \\ I_{n-k}(C(\Gamma)) & \text{if } k = 0, 1, \ldots, n-1 \\ R & \text{if } k \geq n \end{cases}$$

Then

$$\textbf{13.19}\quad \mathfrak{F}_0(\Gamma)(M) \subseteq \mathfrak{F}_1(\Gamma)(M) \subseteq \mathfrak{F}_2(\Gamma)(M) \subseteq \cdots \subseteq \mathfrak{F}_{n-1}(\Gamma)(M) \subseteq \mathfrak{F}_n(\Gamma)\ (M) = R$$

is an increasing sequence of ideals in R. Lemmas 13.4 and 13.9 imply these ideals depend only on our choice of Γ and not on any choice of λ and Δ. We also have $\mathfrak{F}_k(\Gamma)(M) = \mathfrak{F}_k(X\Gamma)(M)$ for all $k \in \mathbb{Z}$ and all $X \in \mathrm{Gl}(n,R)$ by Lemma 13.12.

Let us compute these ideals for Γ and Γ' in Example 13.15. We have

$$\textbf{13.20}\quad \begin{aligned} &\mathfrak{F}_0(\Gamma)(M) = 4\mathbb{Z} && \mathfrak{F}_0(\Gamma')(M) = 4\mathbb{Z} \\ &\mathfrak{F}_1(\Gamma)(M) = 2\mathbb{Z} && \mathfrak{F}_1(\Gamma')(M) = 2\mathbb{Z} \\ &\mathfrak{F}_k(\Gamma)(M) = \mathbb{Z} && \mathfrak{F}_2(\Gamma')(M) = \mathbb{Z} \\ &\text{for all } k \geq 2 && \mathfrak{F}_k(\Gamma')(M) = \mathbb{Z} \quad \text{for all } k \geq 3 \end{aligned}$$

Thus, the sequences $\{\mathfrak{F}_k(\Gamma)(M) \mid k \in \mathbb{Z}\}$ and $\{\mathfrak{F}_k(\Gamma')(M) \mid k \in \mathbb{Z}\}$ are exactly the same. This is always the case. The ideals in 13.19 in fact do not depend on any specific choice of Γ.

Theorem 13.21 Let Γ and Γ' be two R-module bases of M. Then for all $k \in \mathbb{Z}$, $\mathfrak{F}_k(\Gamma)(M) = \mathfrak{F}_k(\Gamma')(M)$.

Proof. In order to prove this result, it suffices to prove the following claim.

Claim. Let $\Gamma = \{m_1, \ldots, m_n\}$ be an R-module basis of M. Let $m \in M$, and set $\Gamma' = \{m_1, \ldots, m_n, m\}$. Then $\mathfrak{F}_k(\Gamma)(M) = \mathfrak{F}_k(\Gamma')(M)$ for all $k \in \mathbb{Z}$.

The theorem follows from repeated applications of the claim. For suppose $\Gamma = \{m_1, \ldots, m_n\}$ and $\Gamma' = \{m'_1, \ldots, m'_p\}$. By the claim,

$$\begin{aligned} \mathfrak{F}_k(\Gamma)(M) &= \mathfrak{F}_k(\Gamma \cup \{m'_1\})(M) = \mathfrak{F}_k(\Gamma \cup \{m'_1, m'_2\})(M) \\ &= \cdots = \mathfrak{F}_k(\Gamma \cup \Gamma')(M) \end{aligned}$$

Similarly, $\mathfrak{F}_k(\Gamma')(M) = \mathfrak{F}_k(\Gamma' \cup \Gamma)(M)$. In particular, $\mathfrak{F}_k(M) = \mathfrak{F}_k(\Gamma')(M)$.

To prove the claim, consider the following diagram of exact sequences:

13.22

$$\begin{array}{ccccccccc} (0) & \longmapsto & K & \xrightarrow{\iota} & R^n & \xrightarrow{f} & M & \longmapsto & (0) \\ & & \theta \downarrow & & \theta \downarrow & & \| & & \\ (0) & \longmapsto & K' & \xrightarrow{\iota} & R^{n+1} & \xrightarrow{\bar{f}} & M & \longmapsto & (0) \end{array}$$

The first row of 13.22 is the short exact sequence determined by Γ and ε. Thus, $f(\sum_{i=1}^n r_i\varepsilon_i) = \sum_{i=1}^n r_i m_i$ and $K = \mathrm{Ker}(f)$. As usual, ε denotes the canonical basis of R^n. The second row of 13.22 is the short exact sequence determined by $\Gamma' = \Gamma \cup \{m\}$. Thus, $\bar{f}(\sum_{i=1}^{n+1} r_i\varepsilon_i') = \sum_{i=1}^n r_i m_i + r_{n+1}m$. Here $\varepsilon' = \{\varepsilon_1', \ldots, \varepsilon_{n+1}'\}$ is the canonical basis of R^{n+1}, and $K' = \mathrm{Ker}(\bar{f})$. The map $\theta : R^n \mapsto R^{n+1}$ is the canonical embedding given by $\theta(\varepsilon_i) = \varepsilon_i'$ for all $i = 1, \ldots, n$. The reader can easily check that $\bar{f}\theta = f$ and $\theta(K) \subseteq K'$.

Since Γ is an R-module basis of M, there exist elements $r_1, \ldots, r_n \in R$ such that $m = \sum_{i=1}^n r_i m_i$. Set $\beta = (\sum_{i=1}^n r_i\varepsilon_i') - \varepsilon_{n+1}'$. Then $\bar{f}(\beta) = 0$, and K' is easily determined. We have

13.23 $K' = \theta(K) \oplus R\beta$.

We have already observed that $\theta(K) + R\beta \subseteq K'$. Since the $n+1$ entry of β is -1, $\theta(K) \cap R\beta = (0)$. Suppose $\delta = \sum_{i=1}^{n+1} s_i\varepsilon_i' \in K'$. Then

$$0 = \bar{f}(\delta) = \sum_{i=1}^n s_i m_i + s_{n+1}m = s_1m_1 + \cdots + s_nm_n + s_{n+1}(r_1m_1 + \cdots + r_nm_n)$$

Therefore, $0 = (s_1 + r_1s_{n+1})m_1 + \cdots + (s_n + r_ns_{n+1})m_n$. This says

$$(s_1 + r_1s_{n+1})\varepsilon_1 + \cdots + (s_n + r_ns_{n+1})\varepsilon_n \in K.$$

Then

$$\delta = \sum_{i=1}^{n+1} s_i\varepsilon_i' = \sum_{i=1}^n (s_i + r_is_{n+1})\varepsilon_i' - s_{n+1}(r_1\varepsilon_1' + \cdots + r_n\varepsilon_n' - \varepsilon_{n+1}') \in \theta(K) + R\beta$$

Thus, $K' \subseteq \theta(K) + R\beta$, and the equality in 13.23 is proved.

Now suppose $\Delta = \{\delta_1, \ldots, \delta_m\}$ is any R-module basis of K. Write $\delta_i = \sum_{j=1}^n c_{ij}\varepsilon_j$ for all $i = 1, \ldots, m$. Then $C(\Gamma,\varepsilon,\Delta) = (c_{ij})$, and $\mathfrak{F}_k(\Gamma)(M) = I_{n-k}((c_{ij}))$. The equality in equation 13.23 implies $\Delta' = \{\theta(\delta_1), \ldots, \theta(\delta_m),\beta\}$ is an R-module basis of K'. Therefore,

13.24 $$C(\Gamma',\varepsilon',\Delta') = \begin{bmatrix} c_{11} & \cdots & c_{1n} & 0 \\ \vdots & & \vdots & \vdots \\ c_{m1} & \cdots & c_{mn} & 0 \\ r_1 & \cdots & r_n & -1 \end{bmatrix} \in M_{(m+1)\times(n+1)}(R)$$

Also, $\mathfrak{F}_k(\Gamma')(M) = I_{n+1-k}(C(\Gamma',\varepsilon',\Delta'))$. Thus, the proof of the claim will be complete once we have the following equality:

13.25 $$I_t(C(\Gamma,\varepsilon,\Delta)) = I_{t+1}(C(\Gamma',\varepsilon',\Delta'))$$

The assertion in equation 13.25 follows easily from the description of $C(\Gamma',\varepsilon',\Delta')$ given in equation 13.24. We leave this point to the exercises at the end of this chapter. ∎

Theorem 13.21 implies the ideals $\{\mathfrak{F}_k(\Gamma)(M) \mid k \in \mathbb{Z}\}$ do not depend on any specific choice of Γ. Hence, we can drop Γ from our notation and write the ideals in 13.19 succinctly as follows:

13.26 $$(0) \subseteq \mathfrak{F}_0(M) \subseteq \mathfrak{F}_1(M) \subseteq \mathfrak{F}_2(M) \subseteq \cdots \subseteq R$$

Definition 13.27 The ideal $\mathfrak{F}_k(M)$ is called the kth Fitting ideal of M.

We can summarize our discussion to this point with the following algorithm.

ALG To compute the kth Fitting ideal of M, proceed as follows:

(a) Construct a short exact sequence of the form

$$(0) \mapsto K \mapsto R^n \mapsto M \mapsto (0)$$

(b) Compute a relations matrix C from the short exact sequence in (a).

(c) $\mathfrak{F}_k(M) = I_{n-k}(C)$.

Step (a) in ALG is done by choosing a basis Γ of M and a free R-module basis λ of R^n (n = the number of generators in Γ) and determining the kernel of the homomorphism $f: R^n \mapsto M$ determined by Γ and λ. Step (b) is done by choosing a basis Δ of $K = \text{Ker}(f)$ and writing these vectors as linear combinations of the free basis λ. As we have seen, the kth Fitting ideal $\mathfrak{F}_k(M)$ will always be the same no matter how we make these choices.

A few observations about this procedure might be helpful. Suppose $(0) \mapsto K \mapsto R^n \mapsto M \mapsto (0)$ is a short exact sequence. Then M has an R-module basis

containing n generators. A relations matrix C obtained from this short exact sequence will be an $m \times n$ matrix in $M_{m\times n}(R)$ for some $m \geq 1$. In particular, the definition of $I_t(C)$ in 4.3 implies the Fitting ideals of M form the following increasing sequence.

13.28 $(0) \subseteq \mathfrak{F}_0(M) \subseteq \mathfrak{F}_1(M) \subseteq \cdots \subseteq \mathfrak{F}_{n-1}(M) \subseteq R$

Also from 4.3, we have $\mathfrak{F}_k(M) = (0)$ whenever $k < n - \min\{m,n\}$.

Suppose M is a free R-module of rank n. Then M has a free R-module basis $\Gamma = \{m_1, \ldots, m_n\}$ containing n elements. M has a presentation of the form

$$(0) \mapsto R^n \xrightarrow{f} M \mapsto (0)$$

The map f sends ε_i to m_i for all $i = 1, \ldots, n$. Since Γ is a free R-module basis of M, f is an isomorphism. Therefore, $K = \operatorname{Ker}(f) = (0)$. Thus, $C = (0, \ldots, 0) \in M_{1\times n}(R)$ is a relation matrix for M. Since $I_t(C) = R$ for any $t \leq 0$ and $I_t(C) = (0)$ for all $t \geq 1$, we have proved the following observation.

13.29 If M is a free R-module of rank n, then

$$\mathfrak{F}_k(M) = \begin{cases} R & \text{if } k \geq n \\ (0) & \text{if } k < n \end{cases}$$

The module in Example 13.15 has the following Fitting ideals: $\mathfrak{F}_k(M) = 0$ for all $k \leq -1$, $\mathfrak{F}_0(M) = 4\mathbb{Z}$, $\mathfrak{F}_1(M) = 2\mathbb{Z}$, and $\mathfrak{F}_k(M) = \mathbb{Z}$ for all $k \geq 2$.

Before proving theorems about Fitting ideals, it will be convenient to introduce some notation. If M is an R-module and n a positive integer, then we will let $M^n = \{(m_1, \ldots, m_n)^t \mid m_i \in M\}$. Thus, M^n is the set of all column vectors of size n with entries from M. M^n is an R-module with addition and scalar multiplication performed componentwise. Clearly, M^n is isomorphic to the direct sum of n copies of M.

If p is another positive integer, the matrices in $M_{p\times n}(R)$ induce R-module homomorphisms from M^n to M^p in the obvious way: If $A = (a_{ij}) \in M_{p\times n}(R)$ and $\xi = (m_1, \ldots, m_n)^t \in M^n$, then we define $A\xi$ by the following equation:

13.30 $$A\xi = \left(\sum_{j=1}^{n} a_{1j}m_j, \sum_{j=1}^{n} a_{2j}m_j, \ldots, \sum_{j=1}^{n} a_{pj}m_j\right)^t$$

Thus, $A\xi$ is just the usual product of a matrix with a column vector, only the column vector has entries from M instead of R. It is easy to check that $B(A\xi) = (BA)\xi$ for all $\xi \in M^n$, $A \in M_{p\times n}(R)$, and $B \in M_{q\times p}(R)$. Also, $I_n\xi = \xi$ for all $\xi \in M^n$.

Our first theorem about Fitting ideals says $\mathfrak{F}_0(M)$ and $\mathrm{Ann}_R(M)$ have the same radical.

Theorem 13.31 Let M be a finitely generated module over a Noetherian ring R. Suppose $\Gamma = \{m_1, \ldots, m_n\}$ is a basis of M. Then $(\mathrm{Ann}_R(M))^n \subseteq \mathfrak{F}_0(M) \subseteq \mathrm{Ann}_R(M)$.

Proof. Let $f: R^n \mapsto M$ be the R-module homomorphism given by $f(\sum_{j=1}^n r_j\varepsilon_j) = \sum_{j=1}^n r_j m_j$. Set $K = \mathrm{Ker}(f)$. Then

$$(0) \mapsto K \overset{\iota}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$$

is a short exact sequence of R-modules. Let $\Delta = \{\delta_1, \ldots, \delta_m\}$ be an R-module basis of K. Since $\mathfrak{F}_0(M)$ does not depend on any specific choice of Δ, we can add more vectors to Δ if need be and assume $m > n$.

Let $\delta_i = \sum_{j=1}^n c_{ij}\varepsilon_j$ for all $i = 1, \ldots, m$. Then $\mathfrak{F}_0(M) = I_n(C)$ where $C = (c_{ij})$. Let $i_1, \ldots, i_n$ be positive integers such that $1 \leq i_1 < \cdots < i_n \leq m$. Set $d = \Delta(i_1, \ldots, i_n; 1, \ldots, n)$, the $n \times n$ minor of C defined by rows $i_1, \ldots, i_n$. Let $\xi = (m_1, \ldots, m_n)^t \in M^n$. Since the rows of C generate all relations among $m_1, \ldots, m_n$, we have $C\xi = \mathrm{O}$. In particular, $(\mathrm{Row}_{i_1}(C); \ldots; \mathrm{Row}_{i_n}(C))\,\xi = \mathrm{O}$. Since $d = \det(\mathrm{Row}_{i_1}(C); \ldots; \mathrm{Row}_{i_n}(C))$,

$$d\xi = (dI_n)\xi = \mathrm{adj}(\mathrm{Row}_{i_1}(C); \ldots; \mathrm{Row}_{i_n}(C))\,(\mathrm{Row}_{i_1}(C); \ldots; \mathrm{Row}_{i_n}(C))\,\xi = \mathrm{O}.$$

Thus, $dm_i = 0$ for all $i = 1, \ldots, n$. Since Γ is a basis of M, we conclude $d \in \mathrm{Ann}_R(M)$. Since $\mathfrak{F}_0(M)$ is generated by all $n \times n$ minors of C, we have $\mathfrak{F}_0(M) \subseteq \mathrm{Ann}_R(M)$.

Conversely, suppose $x_1, \ldots, x_n \in \mathrm{Ann}_R(M)$. Then the row vectors $(x_1,0,\ldots,0)$, $(0,x_2,0,\ldots,0)$, $\ldots$, $(0,\ldots,0,x_n)$ $[\in M_{1\times n}(R)]$ are all relations among the generators $m_1, \ldots, m_n$ of M. We can add these rows to the relation matrix C, forming a new relations matrix C'. Thus,

$$\textbf{13.32}\quad C' = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{m1} & \cdots & c_{mn} \\ x_1 & 0 \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & x_n \end{bmatrix} \in M_{(m+n)\times n}(R)$$

We still have $\mathfrak{F}_0(M) = I_n(C')$. In particular, $x_1x_2 \cdots x_n \in I_n(C') = \mathfrak{F}_0(M)$. Since the x_i are arbitrary elements in $\mathrm{Ann}_R(M)$, we conclude $(\mathrm{Ann}_R(M))^n \subseteq \mathfrak{F}_0(M)$. ∎

Our next result describes how Fitting ideals behave under scalar extension.

Theorem 13.33 Let $g : R \mapsto T$ be a homomorphism of (commutative) Noetherian rings. Let M be a finitely generated R-module. Then $\mathfrak{F}_k(M \otimes_R T) = \mathfrak{F}_k(M)T$ for all $k \in \mathbb{Z}$.

We have seen in Chapter 12 that $M \otimes_R T$ is a finitely generated T-module. Since T is Noetherian, $\mathfrak{F}_k(M \otimes_R T)$ is a well-defined ideal in T for each integer k. The symbols $\mathfrak{F}_k(M)T$ are shorthand for $g(\mathfrak{F}_k(M))T$. Theorem 13.33 says the kth Fitting ideal of the extended module $M \otimes_R T$ is the ideal in T generated by the kth Fitting ideal of M.

Proof of 13.33. Let $\Gamma = \{m_1, \ldots, m_n\}$ be an R-module basis of M. Form the short exact sequence

$$(0) \mapsto K \overset{\iota}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$$

with $f(\varepsilon_j) = m_j$ for all $j = 1, \ldots, n$ and $K = \mathrm{Ker}(f)$. Let $\Delta = \{\delta_1, \ldots, \delta_m\}$ be an R-module basis of K, and set $\delta_i = \sum_{j=1}^n c_{ij}\varepsilon_j$ for all $i = 1, \ldots, m$. Then $\mathfrak{F}_k(M) = I_{n-k}(C)$ where $C = (c_{ij})$.

Theorem 12.23 implies $(*) \otimes_R T$ is a right exact functor. Therefore,

$$\textbf{13.34} \quad K \otimes_R T \overset{\iota\otimes 1}{\mapsto} R^n \otimes_R T \overset{f\otimes 1}{\mapsto} M \otimes_R T \mapsto (0)$$

is a right exact sequence of T-modules. It follows from Theorem 12.20 that $R^n \otimes_R T \cong T^n$ as T-modules. The isomorphism here sends $\varepsilon_i \otimes 1$ to ε_i^T where $\varepsilon^T = \{\varepsilon_1^T, \ldots, \varepsilon_n^T\}$ is the canonical basis of the T-module T^n.

Since the sequence in 13.34 is right exact, $\mathrm{Im}(\iota \otimes 1) = \mathrm{Ker}(f \otimes 1)$. The T-module $\mathrm{Im}(\iota \otimes 1)$ is generated by $\{(\iota \otimes 1)(\delta_1 \otimes 1), \ldots, (\iota \otimes 1)(\delta_m \otimes 1)\}$. The element $(\iota \otimes 1)(\delta_i \otimes 1)$ goes to $\sum_{j=1}^n g(c_{ij})\varepsilon_j^T$ under the isomorphism $R^n \otimes_R T \cong T^n$. Thus, we have the following presentation of the T-module $M \otimes_R T$.

$$\textbf{13.35} \quad \sum_{i=1}^m T\left(\sum_{j=1}^n g(c_{ij})\varepsilon_j^T\right) \mapsto T^n \mapsto M \otimes_R T \mapsto (0)$$

In particular, $(g(c_{ij}))$ is a relations matrix for $M \otimes_R T$. Therefore,

$$\mathfrak{F}_k(M \otimes_R T) = I_{n-k}((g(c_{ij}))) = g(I_{n-k}(C))T = \mathfrak{F}_k(M)T. \quad ∎$$

One important special case of Theorem 13.33 is the following example.

Example 13.36 Suppose S is a multiplicatively closed subset of R. If M is any R-module, we can define an equivalence relation $\sim$ on $M \times S$ by setting $(m,s) \sim (m',s')$ if $s''(s'm - sm') = 0$ for some $s'' \in S$. We will let m/s denote the equivalence class in $M \times S$ which contains (m,s). Set $S^{-1}M = \{m/s \mid (m,s) \in M \times S\}$. The set $S^{-1}M$ becomes an R-module when addition and scalar multiplication are defined as follows:

$$\textbf{13.37} \quad \begin{aligned} (m/s) + (m'/s') &= (s'm + sm') / ss' \\ r(m/s) &= rm/s \end{aligned}$$

If $M = R$, then $S^{-1}R$ becomes a commutative ring with multiplication $(r/s)(r'/s') = rr'/ss'$. $S^{-1}M$ is an $S^{-1}R$-module with scalar multiplication defined by $(r/s)(m/s') = rm/ss'$.

The ring $S^{-1}R$ is called the localization of R at the multiplicative set S. If S is the set of all regular elements of R, that is, $S = R - Z(R)$, then $S^{-1}R$ is just the total quotient ring $Q(R)$ of R. If S is the complement of some prime ideal $\mathfrak{P}$, that is, $S = R - \mathfrak{P}$, then $S^{-1}R$ is $R_{\mathfrak{P}}$, the localization of R at $\mathfrak{P}$ (see Example 10.17).

The $S^{-1}R$-module $S^{-1}M$ is called the localization of M at S. It is easy to see that $M \otimes_R S^{-1}R \cong S^{-1}M$ is $S^{-1}R$-modules. The isomorphism is given by $m \otimes (r/s) \mapsto rm/s$. Hence, Theorem 13.33 implies

$$\textbf{13.38} \quad \mathfrak{F}_k(S^{-1}M) \cong \mathfrak{F}_k(M)S^{-1}R \quad \text{for all } k \in \mathbb{Z}. \qquad \blacksquare$$

Having introduced the necessary notation in equation 13.30, we take this opportunity to present a slight generalization of McCoy's theorem. We will need some of this discussion in our main results on Fitting ideals.

Theorem 13.39 Let $C = (c_{ij}) \in M_{m \times n}(R)$. The homogeneous system of equations $CX = O$ has a nontrivial solution $\xi \in M^n$ if and only if $I_n(C) \subseteq \operatorname{Ann}_R(m)$ for some nonzero $m \in M$.

Proof. We can always add zero rows to C with no loss of generality. Hence, we can assume $m \geq n$. Suppose $CX = O$ has a nontrivial solution $\xi = (m_1, \ldots, m_n)^t \in M^n$. Let $i_1, \ldots, i_n$ be positive integers such that $1 \leq i_1 < \cdots < i_n \leq m$. Set $d = \Delta(i_1, \ldots, i_n; 1, \ldots, n)$, the $n \times n$ minor of C defined by rows $i_1, \ldots, i_n$. Let $C(i_1, \ldots, i_n)$ denote the $n \times n$ submatrix of C determined by rows $i_1, \ldots, i_n$. Equation 2.20 implies $d\xi = \operatorname{adj}(C(i_1, \ldots, i_n))C(i_1, \ldots, i_n)\xi = O$. Since ξ is nonzero, some m_j is nonzero. Then $d\xi = O$ implies $dm_j = 0$. Since d is an arbitrary $n \times n$ minor of C, we have $I_n(C) \subseteq \operatorname{Ann}_R(m_j)$.

Conversely, suppose $I_n(C) \subseteq \mathrm{Ann}_R(m)$ for some nonzero $m \in M$. We proceed by induction on n. If $n = 1$, then $\xi = m \in M^1$ is a nontrivial solution to $CX = O$. Hence we can assume $n > 1$ and pass to the inductive step of the argument. There are two cases to consider here.

First suppose $I_{n-1}(C) \subseteq \mathrm{Ann}_R(m)$. Let C' denote the submatrix of C consisting of the first $n - 1$ columns of C. Thus,

$$C' = \begin{bmatrix} c_{11} & \cdots & c_{1n-1} \\ \vdots & & \vdots \\ c_{m1} & \cdots & c_{mn-1} \end{bmatrix}$$

Then $I_{n-1}(C') \subseteq I_{n-1}(C) \subseteq \mathrm{Ann}_R(m)$. Hence, our induction hypothesis implies there exists a nonzero $\xi' = (m_1, \ldots, m_{n-1})^t \in M^{n-1}$ such that $C'\xi' = O$. Then $\xi = (m_1, \ldots, m_{n-1}, 0)^t \in M^n$ is a nontrivial solution to $CX = O$. Thus, the proof is complete in this case.

Now suppose $I_{n-1}(C)$ is not contained in $\mathrm{Ann}_R(m)$. Then there exists an $(n-1)\times(n-1)$ minor Δ of C such that $\Delta m \neq 0$. Suppose $\Delta = \Delta(i_1, \ldots, i_{n-1}; 1, \ldots, p-1, p+1, \ldots, n)$. Since $m \geq n > n-1$, there is at least one more row, say $\mathrm{Row}_k(C)$, of C other than rows $i_1, \ldots, i_{n-1}$. Let C' be the following submatrix of C:

13.40
$$C' = \begin{bmatrix} c_{i_1 1}, \ldots, c_{i_1 p}, \ldots, c_{i_1 n} \\ \vdots \\ c_{i_u 1}, \ldots, c_{i_u p}, \ldots, c_{i_u n} \\ c_{k1}, \ldots, c_{kp}, \ldots, c_{kn} \\ c_{i_{u+1} 1}, \ldots, c_{i_{u+1} p}, \ldots, c_{i_{u+1} n} \\ \vdots \\ c_{i_{n-1} 1}, \ldots, c_{i_{n-1} p}, \ldots, c_{i_{n-1} n} \end{bmatrix} \in M_{n\times n}(R)$$

In 13.40

$$i_1 < \cdots < i_u < k < i_{u+1} < \cdots < i_{n-1}$$

and $1 \leq p \leq n$. Notice that $\det(C') \in I_n(C)$ and the (k,p)th cofactor of C' is $\pm\,\Delta$. Expand the determinant of C' along the row $(c_{k1}, \ldots, c_{kn})$ using Laplace's expansion. We get $\det(C') = c_{k1}\Delta_1 + \cdots + c_{kn}\Delta_n$ with $\Delta_p m$

$= (\pm\Delta)m \neq 0$. Here Δ_i is the cofactor if c_{ki} in the expansion of $\det(C')$. Since $\Delta_1, \ldots, \Delta_n \in I_{n-1}(C)$, Theorem 2.19 implies $\sum_{j=1}^{n} c_{vj}\Delta_j$ is either zero or a nonzero element in $I_n(C)$ for all $v = 1, \ldots, m$. Since $I_n(C)m = 0$, $\xi = (\Delta_1 m, \Delta_2 m, \ldots, \Delta_n m)^t \in M^n$ is a solution to $CX = O$. Since $\Delta_p m \neq 0$, ξ is a nontrivial solution to $CX = O$. ∎

When $M = R$ in Theorem 13.39, we recover McCoy's theorem. We have the following equivalences. $CX = O$ has a nontrivial solution in $R^n \Leftrightarrow I_n(C) \subseteq \mathrm{Ann}_R(x)$ for some $x \neq 0$ in $R \Leftrightarrow \mathrm{Ann}_R(I_n(C)) \neq (0) \Leftrightarrow rk(C) < n$.

Notice that $m < n$ in Theorem 13.39 implies $I_n(C) = (0)$. Thus, $CX = O$ has a nontrivial solution in M^n for any nonzero R-module M.

It is easy to manufacture an R-module whose relation matrix is a given $m \times n$ matrix in $M_{m\times n}(R)$. We need the following definition.

Definition 13.41 Let $C = (c_{ij}) \in M_{m\times n}(R)$. The R-module $M(C)$ associated to C is defined as follows: $M(C) = R^n/K$ where $K = R\delta_1 + \cdots + R\delta_m$ and $\delta_i = \sum_{j=1}^{n} c_{ij}\varepsilon_j$ for all $i = 1, \ldots, m$.

As usual, $\varepsilon = \{\varepsilon_1, \ldots, \varepsilon_n\}$ is the canonical bases of R^n. If we set $f(\varepsilon_i) = \varepsilon_i + K$ in $M(C)$, then

$$\textbf{13.42}\quad (0) \mapsto K \overset{\iota}{\mapsto} R^n \overset{f}{\mapsto} M(C) \mapsto (0)$$

is a short exact sequence of R-modules. In particular, C is a relations matrix for $M(C)$, and $\mathfrak{F}_k(M(C)) = I_{n-k}(C)$ for all $k \in \mathbb{Z}$.

Definition 13.43 Let M be an R-module. An element $m \in M$ is said to be a torsion element if $rm = 0$, for some regular element $r \in R$. The set of all torsion elements in M will be denoted by $\mathcal{T}(M)$.

The reader will recall that an element $r \in R$ is regular if r is not a zero divisor. Thus, $m \in \mathcal{T}(\mathrm{M})$ if there exists an $r \in R - Z(R)$ such that $rm = 0$. It is easy to check that $\mathcal{T}(M)$ is an R-submodule of M. $\mathcal{T}(M)$ is called the torsion submodule of M. If $\mathcal{T}(M) = (0)$, then M is said to be torsion free. Obviously, $M/\mathcal{T}(M)$ is always a torsion-free R-module. If $\mathcal{T}(M) = M$, then M is called a torsion module.

Example 13.44 Let $\mathfrak{A}$ be an ideal containing a regular element x of R. Then the R-module $M = R/\mathfrak{A}$ is a torsion module since $x(1 + \mathfrak{A}) = 0$. Thus, $\mathcal{T}(R/\mathfrak{A}) = R/\mathfrak{A}$. For example, suppose p is a positive prime in $\mathbb{Z}$. Then the $\mathbb{Z}$-module $\mathbb{Z}/p\mathbb{Z}$ is a torsion module. On the other hand, $\mathbb{Z}/p\mathbb{Z}$ is a field. Thus, the $\mathbb{Z}/p\mathbb{Z}$-module $\mathbb{Z}/p\mathbb{Z}$ is a torsion-free module. ∎

We can use McCoy's theorem to decide when the rows (or columns) of a matrix C are linearly independent and when the module $M(C)$ associated to C is torsion free.

Theorem 13.45 Let $C = (c_{ij}) \in M_{m\times n}(R)$. Set $M = M(C)$. Then
(a) The rows of C are linearly independent in $M_{1\times n}(R)$ if and only if $\mathrm{Ann}_R(I_m(C)) = (0)$.
(b) Suppose $\mathrm{Ann}_R(I_m(C)) = (0)$. Then M is torsion free if and only if $Ra : I_m(C) = Ra$ for all regular elements $a \in R$.

Proof. (a) Partition C into rows: $C = (\delta_1^t; \ldots ;\delta_m^t)$. Here $\delta_i = (c_{i1}, c_{i2}, \ldots, c_{in})^t \in R^n$ for all $i = 1, \ldots, m$. The vectors $\delta_1^t, \ldots, \delta_m^t$ are linearly independent in $M_{1\times n}(R)$ if and only if the homogeneous equation $(x_1, \ldots, x_m)C = \mathrm{O}$ has only the trivial solution $x_1 = 0, \ldots, x_m = 0$. Set $X = (x_1, \ldots, x_m)$. Then $XC = \mathrm{O}$ if and only if $C^tX^t = \mathrm{O}$. Thus, Theorem 13.39 implies the rows of C are linearly independent if and only if $\mathrm{Ann}_R(I_m(C^t)) = (0)$. Since $I_m(C) = I_m(C^t)$, we have that the rows of C are linearly independent if and only if $\mathrm{Ann}_R(I_m(C)) = (0)$. This proves (a).

(b) Suppose $\mathrm{Ann}_R(I_m(C)) = (0)$. Then the rows $\delta_1^t, \ldots, \delta_m^t$ are linearly independent by (a). It easily follows from this that the vectors $\delta_1, \ldots, \delta_m$ are linearly independent in R^n. In particular, 13.42 is a short exact sequence for $M(C)$ with $K = R\delta_1 + \cdots + R\delta_m$ a free R-module of rank m.

Any R-module M is torsion free if whenever $am = 0$ for $m \in M$ and $a \in R - Z(R)$, then $m = 0$. When $M = M(C)$, this statement is equivalent to the following assertion:

13.46 $\mathcal{T}(M(C)) = (0)$ if and only if for every regular element $a \in R$ and for all $(b_1, \ldots, b_n)^t \in R^n$, $a(b_1, \ldots, b_n)^t \in R\delta_1 + \cdots + R\delta_m$ implies $(b_1, \ldots, b_n)^t \in R\delta_1 + \cdots + R\delta_m$.

If $D = (d_{ij}) \in M_{m\times n}(R)$ and $a \in R$, then we will let $\overline{D} = (\overline{d}_{ij})$ denote the image of D in $M_{m\times n}(R/aR)$. Thus, $\overline{D} = (\overline{d}_{ij})$ is the $m\times n$ matrix in $M_{m\times n}(R/aR)$ whose i,jth entry is the coset $d_{ij} + Ra$ in R/aR.

The statement "$a(b_1, \ldots, b_n)^t \in K = R\delta_1 + \cdots + R\delta_m \Rightarrow (b_1, \ldots, b_n)^t \in K$ for any $a \in R - Z(R)$" in 13.46 is equivalent to the following assertion:

13.47 For every regular element $a \in R$, the homogeneous system of equations $(x_1, \ldots, x_m)\overline{C} = \mathrm{O}$ has no nontrivial solution in $M_{1\times m}(R/aR)$.

To see that these two statements are equivalent, first observe that $a(b_1, \ldots, b_n)^t \in R\delta_1 + \cdots + R\delta_m$ if and only if the equation $XC = (ab_1, \ldots, ab_n)$ has a solution $(r_1, \ldots, r_m) \in M_{1\times m}(R)$. Now suppose

$$a(b_1, \ldots, b_n)^t \in K \Rightarrow (b_1, \ldots, b_n)^t \in K$$

for every regular element $a \in R$. Fix $a \in R - Z(R)$. Let $(\bar{r}_1 \ldots, \bar{r}_m) \in M_{1\times m}(R/aR)$ be a solution to $X\overline{C} = O$. Then $(r_1, \ldots, r_m)C = (ab_1, \ldots, ab_n)$ for some $(b_1, \ldots, b_n) \in M_{1\times n}(R)$. Thus, $a(b_1, \ldots, b_n)^t \in R\delta_1 + \cdots + R\delta_m = K$. We conclude that $(b_1, \ldots, b_n)^t \in K$. Therefore,

$$r_1\delta_1^t + \cdots + r_m\delta_m^t = a(b_1, \ldots, b_n) \in a(R\delta_1^t + \cdots + R\delta_m^t)$$

Since $\delta_1^t, \ldots, \delta_m^t$ are linearly independent over R, we have $r_i \in Ra$ for all $i = 1, \ldots, m$. Thus, $\bar{r}_1 = \cdots = \bar{r}_m = 0$ in R/aR, and $X\overline{C} = O$ has only the trivial solution.

The fact that 13.47 implies the statement "$a(b_1, \ldots, b_n)^t \in K \Rightarrow (b_1, \ldots, b_n)^t \in K$ for all regular $a \in R$" is easy. We leave this to the reader.

It now follows that $M(C)$ being torsion free is equivalent to the statement in 13.47. By (a), 13.47 is equivalent to $\mathrm{Ann}_{R/aR}(I_m(\overline{C})) = (0)$ for all regular $a \in R$. $\mathrm{Ann}_{R/aR}(I_m(\overline{C})) = (0)$ if and only if $Ra : I_m(C) = Ra$. This completes the proof of (b). ∎

Suppose $C \in M_{m\times n}(R)$ has more rows than columns, that is, $m > n$. Then $I_m(C) = (0)$, and $\mathrm{Ann}_R(I_m(C)) = R$. Theorem 13.45a then implies the rows of C are linearly dependent. This is certainly a familiar result if R is a field.

Let us now return to Fitting ideals. One important use of these ideals is to decide when a right exact sequence is actually exact. Suppose we have the following presentation of M:

13.48 $R^p \overset{g}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$

Then $\mathrm{Im}(g) = \mathrm{Ker}(f)$, and $\mathrm{Im}(f) = M$. We say the sequence in 13.48 is exact if g is injective. Thus, the right exact sequence given in 13.48 is exact if

$$(0) \to R^p \overset{g}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$$

is a short exact sequence. It follows from Theorem 5.10 that if $p > n$, then g is never injective. Hence, we assume $p \leq n$. Write $p = n - r$. Then $0 \leq r < n$. We then have the following theorem.

Theorem 13.49 Suppose

$$(*) : R^{n-r} \overset{g}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$$

is a right exact sequence of modules over the Noetherian ring R. $(*)$ is a short exact sequence if and only if $\mathfrak{F}_r(M)$ contains a regular element of R.

Proof. Let $\varepsilon' = \{\varepsilon'_1, \ldots, \varepsilon'_{n-r}\}$ denote the canonical basis of R^{n-r} and $\varepsilon = \{\varepsilon_1, \ldots, \varepsilon_n\}$ the canonical basis of R^n. Set $\delta_i = g(\varepsilon'_i)$ for $i = 1, \ldots, n - r$. Let $K = R\delta_1 + \cdots + R\delta_{n-r}$. Since $(*)$ is right exact,

13.50 $(0) \mapsto K \overset{\iota}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$

is a short exact sequence.

Let $\delta_i = \sum_{j=1}^{n} c_{ij}\varepsilon_j$ for $i = 1, \ldots, n - r$, and set $C = (c_{ij}) \in M_{(n-r)\times n}(R)$. We have seen in equation 10.14 that the map g in $(*)$ is given by $g(\lambda) = C^t\lambda$ for all $\lambda \in R^{n-r}$. It follows from Theorem 5.36b that g is injective if and only if $\mathrm{Ann}_R(I_{n-r}(C^t)) = (0)$. Since C is a relations matrix of M, we have $I_{n-r}(C^t) = I_{n-r}(C) = \mathfrak{F}_r(M)$. Thus, g is injective if and only if $\mathrm{Ann}_R(\mathfrak{F}_r(M)) = (0)$. In other words, $(*)$ is a short exact sequence if and only if $\mathrm{Ann}_R(\mathfrak{F}_r(M)) = (0)$. The theorem is then a consequence of the following general fact about Noetherian rings.

13.51 Let $\mathfrak{A}$ be an ideal in a Noetherian ring R. Then $\mathrm{Ann}_R(\mathfrak{A}) = (0)$ if and only if $\mathfrak{A}$ contains a regular element of R.

If $\mathfrak{A}$ contains a regular element of R, then clearly $\mathrm{Ann}_R(\mathfrak{A}) = (0)$. Suppose $\mathrm{Ann}_R(\mathfrak{A}) = (0)$. Notice $\mathrm{Ann}_R(\mathfrak{A}) = (0) : \mathfrak{A}$. Thus, $(0) : \mathfrak{A} = (0)$. It follows from Theorem 11.16 that $\mathfrak{A}$ is contained in no associated prime of (0). Suppose these primes are $\mathfrak{P}_1, \ldots, \mathfrak{P}_d$. Then Lemma 11.10b implies $\mathfrak{A}$ is not contained in $\bigcup_{i=1}^{d} \mathfrak{P}_i$. By Theorem 11.13, $Z(R) = \bigcup_{i=1}^{d} \mathfrak{P}_i$. Therefore, $\mathfrak{A}$ contains a regular element of R. This completes the proof of 13.51 and, consequently, the proof of the theorem. ∎

Using the same notation as in Theorem 13.49, notice that $\mathfrak{F}_r(M)$ is the smallest Fitting ideal of M which can possibly be nonzero. Since $(*)$ is a presentation of M, M has an $(n - r) \times n$ relations matrix C. If $j \in \{0, \ldots, r - 1\}$, then $\mathfrak{F}_j(M) = I_{n-j}(C) = (0)$ since $n - j > n - r = \min\{n - r, n\}$.

The results we have presented in this chapter can be found in many places in the literature. One good source for all this material is [7]. We will finish this chapter with one more important theorem from this paper. We need a couple of definitions from homological algebra.

An R-module M is said to be projective if M is a direct summand of a free R-module F. Thus, M is projective if $M \oplus N = F$ for some R-module N and some free R-module F. Notice that any free R-module is projective. Almost all modern, graduate level texts in abstract algebra cover the basic facts about projective modules. A good reference is [6].

Definition 13.52 A R-module M is said to have projective dimension ≤ 1, if whenever $(0) \mapsto P_1 \mapsto P_0 \mapsto M \mapsto (0)$ is a short exact sequence of R-

modules with P_0 projective, then P_1 is also projective. If M has projective dimension ≤ 1, we will write $pd_R(M) \leq 1$.

Suppose M is a finitely generated module over a Noetherian ring R. Then M admits a short exact sequence like that in 13.1. Suppose we also assume $pd_R(M) \leq 1$. Then K is a projective R-module. Since R is Noetherian, K is also finitely generated. It follows from this that K is a direct summand of a finitely generated free R-module. (See the exercises at the end of this chapter.) If we also assume R is a local ring, then K is a free R-module by Corollary 10.22. Since $K \subseteq R^n$, the rank of K is $n - r$ for some $r \geq 0$. Thus, if R is a local ring and $pd_R(M) \leq 1$, then M admits a short exact sequence of the form

$$(*) : (0) \mapsto R^{n-r} \mapsto R^n \mapsto M \mapsto (0)$$

The converse of this statement is also true. If there is a short exact sequence like $(*)$, then $pd_R(M) \leq 1$. A proof of this fact can be found in [6, Thm. 6.19]. Hence, for a (Noetherian) local ring R, $pd_R(M) \leq 1$ is equivalent to M has a short exact sequence of the form $(*)$.

Theorem 13.49 gives us a way to manufacture modules M for which $pd_R(M) \leq 1$. If

$$(*) : R^{n-r} \mapsto R^n \mapsto M \mapsto (0)$$

is a right exact sequence for which $\mathfrak{F}_r(M)$ contains a regular element, then $pd_R(M) \leq 1$. For Theorem 13.49 implies

$$(0) \mapsto R^{n-r} \mapsto R^n \mapsto M \mapsto (0) \quad \text{is exact}$$

Thus, $pd_R(M) \leq 1$ by the discussion above.

We can now prove the following theorem.

Theorem 13.53 (J. Lipman) Suppose (R,m,k) is a Noetherian local ring. Let M be a finitely generated R-module. Let $r \geq 0$. Then the following statements are equivalent:

(a) The smallest nonzero Fitting ideal of M is $\mathfrak{F}_r(M)$, and $\mathfrak{F}_r(M) = Rx$ for some regular element $x \in R$.

(b) $pd_R(M) \leq 1$, and $M/\mathcal{T}(M)$ is a free R-module of rank r.

Proof. (b) $\Rightarrow$ (a): If $M/\mathcal{T}(M)$ is a free R-module of rank r, then $M/\mathcal{T}(M) \cong R^r$, and $\mathcal{T}(M) \oplus R^r \cong M$. We can assume with no lost of generality that $M = \mathcal{T}(M) \oplus R^r$. We first show

13.54 $\mathfrak{F}_p(M) = \mathfrak{F}_{p-r}(\mathcal{T}(M))$ for all $p \geq 0$.

Since $\mathcal{T}(M)$ is a submodule of the finitely generated R-module M and R is Noetherian, $\mathcal{T}(M)$ is a finitely generated R-module. Hence, $\mathcal{T}(M)$ admits a short exact sequence of the following form:

$$\textbf{13.55}\quad (0) \mapsto K \overset{\iota}{\mapsto} R^n \overset{f}{\mapsto} \mathcal{T}(M) \mapsto (0)$$

Suppose $K = R\delta_1 + \cdots + R\delta_m$ with $\delta_i = \sum_{j=1}^{n} c_{ij}\varepsilon_j$ for all $i = 1, \ldots, m$. Then $C = (c_{ij}) \in M_{m\times n}(R)$ is a relations matrix for $\mathcal{T}(M)$. We can imbed the presentation of $\mathcal{T}(M)$ given in 13.55 into a short exact sequence for M in the following way:

$$\begin{array}{ccccccccc} \textbf{13.56}\quad (0) & \mapsto & K & \overset{\iota}{\mapsto} & R^n & \overset{f}{\mapsto} & \mathcal{T}(M) & \mapsto & (0) \\ & & \theta\downarrow & & \downarrow\theta & & \downarrow\iota & & \\ & (0) & \mapsto & K' & \overset{\iota}{\mapsto} & R^{n+1} & \overset{f'}{\mapsto} & M & \mapsto (0) \end{array}$$

In the diagram in 13.56, ι denotes the inlusion map. θ is given by $\theta(\varepsilon_i) = \varepsilon_i'$ for $i = 1, \ldots, n$ where $\varepsilon' = \{\varepsilon_1', \ldots, \varepsilon_{n+r}'\}$ is the canonical basis of R^{n+r} and $\varepsilon = \{\varepsilon_1, \ldots, \varepsilon_n\}$ is the canonical basis of R^n. The map f' is given by $f'(\varepsilon_i') = f(\varepsilon_i)$ if $i = 1, \ldots, n$ and

$$f'(\varepsilon_i') = (0, \ldots, \underset{j}{1}, \ldots, 0)^t \in R^r$$

if $i = n + j$ with $j = 1, \ldots, r$.

The j below the vector $(0, \ldots, 1, \ldots, 0)^t$ indicates that 1 lies in the jth entry of $f'(\varepsilon_i')$. The map f' makes sense because $M = \mathcal{T}(M) \oplus R^r$. $K' = \operatorname{Ker}(f')$. It is easy to check that 13.56 is a commutative diagram of short exact sequences in which $\theta(K) = K'$. In particular,

$$\textbf{13.57}\quad C' = \begin{bmatrix} c_{11} & \cdots & c_{1n} & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ c_{m1} & \cdots & c_{mn} & 0 & \cdots & 0 \end{bmatrix} \in M_{m\times(n+r)}(R)$$

is a relations matrix for M. Thus, for all $p \geq 0$,

$$\mathfrak{F}_p(M) = I_{(n+r)-p}(C') = I_{(n+r)-p}(C) = \mathfrak{F}_{p-r}(\mathcal{T}(M))$$

This proves 13.54.

Now every element in $\mathcal{T}(M)$ is torsion, and $\mathcal{T}(M)$ is a finitely generated R-module. Hence, $a\mathcal{T}(M) = (0)$ for some regular element $a \in R$. In other words, $\operatorname{Ann}_R(\mathcal{T}(M))$ contains a regular element a of R. It follows from Theo-

rem 13.31 that $\mathfrak{F}_0(\mathcal{T}(M))$ contains a regular element of R. Equation 13.54 implies $\mathfrak{F}_r(M)$ contains a regular element of R. If $p < r$, again by 13.54, $\mathfrak{F}_p(M) = \mathfrak{F}_{p-r}(\mathcal{T}(M)) = (0)$ since $p - r < 0$. Therefore, $\mathfrak{F}_r(M)$ is the smallest, nonzero Fitting ideal of M.

Since $M = \mathcal{T}(M) \oplus R^r$, $pd_R(\mathcal{T}(M)) \leq pd_R(M)$. This follows from [6, p. 377]. Thus, $pd_R(\mathcal{T}(M)) \leq 1$. Since (R,m,k) is a local ring, $\mathcal{T}(M)$ admits a short exact sequence of the following form:

13.58 $(0) \mapsto R^s \mapsto R^n \mapsto \mathcal{T}(M) \mapsto (0)$

In 13.58, notice that $s \leq n$. This follows from Theorem 5.10. Suppose $s < n$. Then $\mathfrak{F}_0(\mathcal{T}(M)) = I_n(\overline{C}) = (0)$. Here $\overline{C}$ is any $s \times n$ relations matrix coming from 13.58. But $\mathfrak{F}_0(\mathcal{T}(M)) = \mathfrak{F}_r(M) \neq (0)$. Therefore, $s = n$. In particular, $\mathfrak{F}_r(M) = \mathfrak{F}_0(\mathcal{T}(M)) = I_n(\overline{C}) = Rx$ where $x = \det(\overline{C})$. Since the ideal $\mathfrak{F}_r(M)$ contains a regular element of R, x must be regular. This completes the proof of (a).

(a) $\Rightarrow$ (b): Let $\Gamma = \{m_1, \ldots, m_n\}$ be an R-module basis of M. Let

13.59 $(0) \mapsto K \overset{\iota}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$

be the corresponding short exact sequence of M. Thus, $f(\varepsilon_j) = m_j$ for all $j = 1, \ldots, n$ and $K = \mathrm{Ker}(f)$. Suppose $K = R\delta_1 + \cdots + R\delta_m$ with $\delta_i = \sum_{j=1}^{n} c_{ij}\varepsilon_j$ for each $i = 1, \ldots, m$. Then $C = (c_{ij}) \in M_{m \times n}(R)$ is a relations matrix of M, and $\mathfrak{F}_r(M) = I_{n-r}(C)$. We are assuming that $I_{n-r}(C) = \mathfrak{F}_r(M) = Rx$ for some regular element $x \in R$.

The ideal $I_{n-r}(C)$ is generated by all $(n - r) \times (n - r)$ minors of C. Since R is a local ring, Corollary 10.21 implies $\mathfrak{F}_r(M)$ is generated by one of these minors. By permuting the ε_j and δ_i if need be, we can assume the minor $\Delta = \Delta(1, \ldots, n - r; 1, \ldots, n - r)$ generates $\mathfrak{F}_r(M)$. Since $R\Delta = \mathfrak{F}_r(M)$ contains the regular element x, Δ itself must be regular. We can replace x with Δ and assume $\mathfrak{F}_r(M) = Rx$ with

$$x = \det \begin{bmatrix} c_{11}, & \ldots, & c_{1n-r} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ c_{n-r,1}, & \ldots, & c_{n-r,n-r} \end{bmatrix}$$

a regular element of R.

Let C' be the following $(n - r) \times n$ submatrix of C:

$$\textbf{13.60}\quad C' = \begin{bmatrix} c_{11} & \cdots & c_{1,n-r} & c_{1,n-r+1} & \cdots & c_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ c_{n-r,1} & \cdots & c_{n-r,n-r} & c_{n-r,n-r+1} & \cdots & c_{n-r,n} \end{bmatrix}$$

Set

$$D = \begin{bmatrix} c_{11} & \cdots & c_{1,n-r} \\ \vdots & & \vdots \\ c_{n-r,1} & \cdots & c_{n-r,n-r} \end{bmatrix}$$

Let $h \in \{1, \ldots, n-r\}$. For each $i = 1, \ldots, n-r$, let H_{ih} denote the (i,h)th cofactor of D. From Laplace's expansion (Theorem 2.19b), we have

$$\textbf{13.61}\quad \sum_{i=1}^{n-r} c_{ih}H_{ih} = x$$

and

$$\sum_{i=1}^{n-r} c_{ik}H_{ih} = 0 \quad \text{for } 1 \le k \ne h \le n-r$$

Each row of C' determines a relation among the generators $m_1, \ldots, m_n$. Thus, $\sum_{j=1}^{n} c_{ij}m_j = 0$ for $i = 1, \ldots, n-r$. In particular,

$$H_{1h}\sum_{j=1}^{n} c_{1j}m_j = \cdots H_{n-r,h}\sum_{j=1}^{n} c_{n-r,j}m_j = 0$$

Adding these equations together and using the relations in 13.61, we have

$$0 = \sum_{i=1}^{n-r} H_{ih}\left(\sum_{j=1}^{n} c_{ij}m_j\right) = \sum_{j=1}^{n}\left(\sum_{i=1}^{n-r} c_{ij}H_{ih}\right)m_j$$

$$= xm_h + \sum_{j=n-r+1}^{n}\left(\sum_{i=1}^{n-r} c_{ij}H_{ih}\right)m_j$$

Set $d_j = \sum_{i=1}^{n-r} c_{ij}H_{ih}$ for $j = n - r + 1, \ldots, n$. Then $xm_h + \sum_{j=n-r+1}^{n} d_j m_j = 0$ and

$$d_{n-r+1}, \ldots, d_n \in I_{n-r}(C') \subseteq I_{n-r}(C) = \mathfrak{F}_r(M) = Rx$$

Thus, each d_j is divisible by x. Therefore, $d_j/x \in R$ for each $j = n - r + 1, \ldots, n$. We now have $x(m_h + \sum_{j=n-r+1}^{n} (d_j/x)m_j) = 0$. Since x is a regular element of R, we have the following assertion: For each $h = 1, \ldots, n - r$, there exist $d_{n-r+1}, \ldots, d_n \in Rx$ such that

$$\textbf{13.62} \quad m_h + \sum_{j=n-r+1}^{n} \left(\frac{d_j}{x}\right) m_j \in \mathcal{T}(M)$$

Now let $\overline{m}_1, \ldots, \overline{m}_n$ denote the images of $m_1, \ldots, m_n$, respectively, in $M/\mathcal{T}(M)$. Then the assertion in 13.62 implies $M/\mathcal{T}(M) = R\overline{m}_{n-r+1} + \cdots R\overline{m}_n$. Thus, $\Gamma_1 = \{\overline{m}_{n-r+1}, \ldots, \overline{m}_n\}$ is an R-module basis of $M/\mathcal{T}(M)$. We will show Γ_1 is a free R-module basis of $M/\mathcal{T}(M)$.

Let Q denote the total qoutient ring of R. Thus, $Q = S^{-1}R$ where S is the set of all regular elements in R (see Example 13.36). Since $\mathcal{T}(M)$ is a torsion module, $\mathcal{T}(M) \otimes_R Q = (0)$. Since $(0) \mapsto \mathcal{T}(M) \mapsto M \mapsto M/\mathcal{T}(M) \mapsto (0)$ is an exact sequence of R-modules, Theorem 12.23 implies $M \otimes_R Q \cong (M/\mathcal{T}(M)) \otimes_R Q$. Since $M/\mathcal{T}(M)$ is a torsion-free R-module, it is easy to see the map $M/\mathcal{T}(M) \to (M/\mathcal{T}(M)) \otimes_R Q$ given by $z \mapsto z \otimes 1$ is an injective R-module homomorphism. Hence, $M/\mathcal{T}(M)$ can be identified with an R-submodule of $M \otimes_R Q$. In particular, Γ_1 is a Q-module basis of $M \otimes_R Q$.

By Theorem 13.33, $\mathfrak{F}_p(M \otimes_R Q) = \mathfrak{F}_p(M)Q$ for all $p \in \mathbb{Z}$. Thus, $\mathfrak{F}_r(M \otimes_R Q) = Qx = Q$. This last equality comes from the fact that x regular in R implies x is a unit in Q. Since we are assuming $\mathfrak{F}_r(M)$ is the smallest nonzero Fitting ideal of M, we also have $\mathfrak{F}_p(M \otimes_R Q) = (0)$ for all $p < r$.

Now suppose $\sum_{j=n-r+1}^{n} s_j\overline{m}_j = 0$ in $M/\mathcal{T}(M) \subseteq M \otimes_R Q$. Here $s_{n-r+1}, \ldots, s_n \in R$. Since $M \otimes_R Q$ is generated as a Q-module by Γ_1, we have the following short exact sequence:

$$\textbf{13.63} \quad (0) \mapsto \mathrm{Ker}(\mu) \mapsto Q^r \xrightarrow{\mu} M \otimes_R Q \mapsto (0)$$

In 13.63, μ is the Q-module homorphism given by $\mu(\varepsilon_i) = \overline{m}_{n-r+i} \otimes 1$ for all $i = 1, \ldots, r$. The r-tuple $(s_{n-r+1}, \ldots, s_n)^t$ lies in $\mathrm{Ker}(\mu)$ and hence appears as a row in some suitably chosen relations matrix E of $M \otimes_R Q$. Therefore, for all $j = n - r + 1, \ldots, n$, $s_j \in \mathfrak{F}_{r-1}(M \otimes_R Q) = (0)$. We have now shown that $\overline{m}_{n-r+1}, \ldots, \overline{m}_n$ are linearly independent over R. Thus, Γ_1 is a free R-module basis of $M/\mathcal{T}(M)$.

We have now shown that $M/\mathcal{T}(M)$ is a free R-module of rank r. It remains to show $pd_R(M) \leq 1$. We again consider the submatrix C' of C given by equation

13.60. We first note that the rows of C' are linearly independent. For suppose $(x_1, \ldots, x_{n-r})C' = O$. Then $(x_1, \ldots, x_{n-r})D = O$. Since $\det(D) = x$, a regular element of R, $x_1 = \cdots = x_{n-r} = 0$ by Theorem 13.45a. Thus, the rows of C' are linearly independent.

Suppose $\sum_{j=1}^{n} a_j m_j = 0$ for some elements $a_1, \ldots, a_n \in R$. Thus, $(a_1, \ldots, a_n)$ is a relation among the generators $m_1, \ldots, m_n$ of M. Let $h \in \{1, \ldots, n\}$. Consider the following $(n - r + 1) \times (n - r + 1)$ matrix:

13.64 $$C_h = \left[\begin{array}{c|c} c_{1h} & \\ \vdots & D \\ c_{n-r,h} & \\ \hline a_h & a_1, \ldots, a_{n-r} \end{array}\right]$$

Since $(a_1, \ldots, a_n)$ is a relation among $m_1, \ldots, m_n$, we can add $(a_1, \ldots, a_n)$ as a row to the original relations matrix C. Then $\det(C_h) = 0$ for $h = 1, \ldots, n - r$ and $\det(C_h) \in I_{n-r+1}(C) = \mathfrak{F}_{r-1}(M) = (0)$ for $h = n - r + 1, \ldots, n$. Thus, $\det(C_h) = 0$ for all $h = 1, \ldots, n$. Expand $\det(C_h)$ using Laplace's theorem along the first column (from the bottom up). We get $0 = xa_h + \sum_{i=1}^{n-r} c_{ih} d_i$ with $d_1, \ldots, d_{n-r} \in I_{n-r}(C) = \mathfrak{F}_r(M) = Rx$. Since x is a regular element of R, we have $a_h + \sum_{i=1}^{n-r} c_{ih}(d_i/x) = 0$ for $h = 1, \ldots, n$. Notice also that the elements $d_i/x \in R$ are the same for every h. Set $x_i = d_i/x$ for $i = 1, \ldots, n - r$. Then $a_h + \sum_{i=1}^{n-r} x_i c_{ih} = 0$ for all $h = 1, \ldots, n$. Thus, $(a_1, \ldots, a_n) \in RS(C')$.

Now $(a_1, \ldots, a_n)$ is an arbitrary relation among the generators in Γ. Since $(a_1, \ldots, a_n) \in RS(C')$, we conclude C' is a relations matrix for M. In particular, M admits a short exact sequence of the form

13.65 $$(0) \mapsto \sum_{i=1}^{n-r} R\delta_i \mapsto R^n \mapsto M \mapsto (0)$$

Here $\delta_i = \mathrm{Row}_i(C')^t$ for $i = 1, \ldots, n - r$. We have noted that the rows of C' are linearly independent over R. Thus, $\sum_{i=1}^{n-r} R\delta_i$ is a free R-module with free R-module basis $\{\delta_1, \ldots, \delta_{n-r}\}$. It follows from [6, Thm.6.19] that $pd_R(M) \leq 1$. This completes the proof of (b). ∎

We conclude this chapter with one obvious refinement of Theorem 13.53.

Corollary 13.66 Let (R,m,k) be a Noetherian, local ring. Let M be a finitely generated R-module. Suppose $\mathfrak{F}_r(M)$ is the smallest nonzero Fitting ideal of M. Then M is a free R-module if and only if $\mathfrak{F}_r(M) = R$.

Proof. This result follows immediately from equation 13.29 if M is a free R-module. So, suppose the smallest, nonzero Fitting ideal of M is $\mathfrak{F}_r(M)$ and $\mathfrak{F}_r(M) = R$. Theorem 13.53 implies $M/\mathcal{T}(M)$ is a free R-module of rank r. It follows from 13.54 that $\mathfrak{F}_0(\mathcal{T}(M)) = \mathfrak{F}_r(M) = R$. We have also observed in the proof of Theorem 13.53 that $\text{Ann}_R(\mathcal{T}(M)) \neq (0)$. Theorem 13.31 implies $R = \mathfrak{F}_0(\mathcal{T}(M)) \subseteq \text{Ann}_R(\mathcal{T}(M))$. Therefore, $\mathcal{T}(M) = (0)$, and M is a free R-module of rank r. ∎

There is a slight generalization of Corollary 13.66 for projective modules of rank r. See Exercise 18 at the end of this chapter. The ideas in this chapter can also be used to prove the Hilbert-Burch theorem, an important result in both commutative ring theory and algebraic geometry. The reader is referred to Appendix D at the end of this book for a proof of this theorem.

EXERCISES

1. Let $\Gamma = \{m_1, \ldots, m_n\}$ be an R-module basis of M. Let $X = (x_{ij}) \in \text{Gl}(n,R)$. Show

$$X\Gamma = \{m_1', \ldots, m_n'\} \qquad (m_i' = \sum_{j=1}^{n} x_{ij}m_j)$$

   is an R-module basis of M. Show Γ is a free R-module basis of M if and only if $X\Gamma$ is a free R-module basis of M.
2. Show $\Delta' = \{2\varepsilon_1, 2\varepsilon_2, \varepsilon_1 + \varepsilon_2 + \varepsilon_3\}$ is an R-module basis of K' in Example 13.15.
3. Prove 13.25.
4. Compute the Fitting ideals of each $\mathbb{Z}$-module listed in Exercise 18 of Chapter 10.
5. Compute the Fitting ideals for $\mathfrak{A} = (X,Y)$ in Exercise 17 of Chapter 10.
6. Let $A \in M_{n\times n}(F)$ and $M = F^n$. Here F is a field. Show M is a finitely generated $F[X]$-module via $f(X)m = f(A)(m)$. Compute the Fitting ideals of M.
7. Show $S^{-1}M \cong M \otimes_R S^{-1}R$ for any R-module M and any multiplicatively closed subset S of R.
8. Let M be an R-module. Show $\mathcal{T}(M)$ is a submodule of M and $M/\mathcal{T}(M)$ is torsion free.
9. Exhibit a torsion-free R-module which is not a free R-module.
10. Show 13.47 implies 13.46.

11. Suppose K is a projective R-module. Let $(0) \mapsto M \mapsto N \mapsto K \mapsto (0)$ be a short exact sequence of R-modules. Show $N \cong M \oplus K$.
12. Use Exercise 11 to show the following: If K is a finitely generated R-module which is projective, then K is a direct summand of R^n for some integer n.
13. Suppose M is a finitely generated R-module. If $M = \mathfrak{T}(M)$, show $\text{Ann}_R(M)$ contains a regular element of R.
14. Show $\theta(K) = K'$ in 13.56.
15. Suppose $M = M_1 \oplus M_2$ with M_1 and M_2 finitely generated R-modules. We assume R is Noetherian. Show

$$\mathfrak{F}_k(M) = \sum_{i+j=k} \mathfrak{F}_i(M_1)\, \mathfrak{F}_j(M_2) \quad \text{for all } k \in \mathbb{Z}$$

16. Prove the following theorem of R. E. MacRae: Suppose $(0) \mapsto M \mapsto N \mapsto P \mapsto (0)$ is a short exact sequence of finitely generated modules over a Noetherian ring R. Then $\mathfrak{F}_k(M)\mathfrak{F}_k(P) \subseteq \mathfrak{F}_k(N)$. (Hint: See [8].)
17. Let M be a finitely generated module over a Noetherian, local ring (R,m,k). Set $\mu = \dim_k(M/mM)$. Show the following statements are equivalent:
    (a) $\mu = r$.
    (b) $\mathfrak{F}_r(M) = R$ and $\mathfrak{F}_{r-1}(M) \subseteq m$.
18. Let M be a finitely generated, projective R-module. M is said to have rank r if $\dim_{k(\mathfrak{P})}(M_{\mathfrak{P}}/\mathfrak{P}M_{\mathfrak{P}}) = r$ for all prime ideals $\mathfrak{P}$ of R. The notation here is as follows: $M_{\mathfrak{P}} = S^{-1}M$ where $S = R - \mathfrak{P}$. $k(\mathfrak{P})$ is the field $Q(R/\mathfrak{P})$ Show the following statements are equivalent:
    (a) M is projective of rank r.
    (b) $\mathfrak{F}_0(M) = \cdots = \mathfrak{F}_{r-1}(M) = (0)$, and $\mathfrak{F}_r(M) = R$.
19. Let S be a multiplicatively closed subset of R. Suppose

$$(0) \mapsto M \overset{f}{\mapsto} N \overset{g}{\mapsto} P \mapsto (0)$$

is a short exact sequence of R-modules. Show

$$(0) \mapsto M \otimes_R S^{-1}R \overset{f\otimes 1}{\mapsto} N \otimes_R S^{-1}R \overset{g\otimes 1}{\mapsto} P \otimes_R S^{-1}R \mapsto (0)$$

is an exact sequence of $S^{-1}R$-modules. (Hint: Use Exercise 7.)

14
Principal Ideal Rings

In this chapter, we will develop a structure theorem for principal ideal rings. We will use this theorem in the next chapter when discussing the Smith normal form of a matrix. The reader will recall that a commutative ring R is called a principal ideal ring if every ideal of R is principal. If R is a principal ideal ring, we will abbreviate this statement by saying R is a PIR. Thus, a commutative ring R is a PIR if every ideal $\mathfrak{A}$ of R is principal, that is, $\mathfrak{A} = Rx$ for some $x \in \mathfrak{A}$. If R is a PIR and an integral domain, then R is called a principal ideal domain (abbreviated PID). Certainly, any homomorphic image of a PID is a PIR. Thus, the rings given in 1.7 together with all their homomorphic images are PIRs. If $R_1, \ldots, R_n$ are PIRs, then a simple argument shows $R_1 \oplus \cdots \oplus R_n$ is also a PIR. Thus, for example, $\mathbb{Z}/n_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/n_r\mathbb{Z}$ is a PIR for any choice of integers $n_1, \ldots, n_r$.

We first need a version of the Chinese Remainder Theorem.

14.1 Suppose $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ are ideals of a commutative ring R such that $\mathfrak{A}_i + \mathfrak{A}_j = R$ whenever $i \neq j$. Then $R/\bigcap_{i=1}^n \mathfrak{A}_i \cong \bigoplus_{i=1}^n R/\mathfrak{A}_i$.

Proof. The map $r \mapsto (r + \mathfrak{A}_1, \ldots, r + \mathfrak{A}_n)$ is clearly a ring homomorphism from R to $R/\mathfrak{A}_1 \oplus \cdots \oplus R/\mathfrak{A}_n$. The kernel of this map is $\bigcap_{i=1}^n \mathfrak{A}_i$. Hence, we have a monomorphism $\theta : R/\bigcap_{i=1}^n \mathfrak{A}_i \mapsto R/\mathfrak{A}_1 \oplus \cdots \oplus R/\mathfrak{A}_n$ given by $\theta(r + \bigcap_{i=1}^n \mathfrak{A}_i) = (r + \mathfrak{A}_1, \ldots, r + \mathfrak{A}_n)$. We claim the map θ is surjective. This is an immediate consequence of the following assertion:

14.2 If $b_1, \ldots, b_n \in R$, then there exists an $a \in R$ such that $a - b_i \in \mathfrak{A}_i$, that is, $a \equiv b_i \bmod \mathfrak{A}_i$ for all $i = 1, \ldots, n$.

We prove 14.2 by induction on n. Suppose $n = 2$. Since $\mathfrak{A}_1 + \mathfrak{A}_2 = R$, $1 = a_1 + a_2$ with $a_1 \in \mathfrak{A}_1$ and $a_2 \in \mathfrak{A}_2$. Set $a = b_1a_2 + b_2a_1$. Then $a \equiv b_i \bmod \mathfrak{A}_i$ for $i = 1,2$.

For the inductive step, first notice that

$$R = \prod_{i=1}^{n-1} (\mathfrak{A}_i + \mathfrak{A}_n) \subseteq \mathfrak{A}_n + \mathfrak{A}_1\mathfrak{A}_2 \cdots \mathfrak{A}_{n-1} \subseteq R$$

Therefore, $\mathfrak{A}_n + \mathfrak{A}_1\mathfrak{A}_2 \cdots \mathfrak{A}_{n-1} = R$. Since $\mathfrak{A}_1\mathfrak{A}_2 \cdots \mathfrak{A}_{n-1} \subseteq \cap_{i=1}^{n-1} \mathfrak{A}_i$, $\mathfrak{A}_n + (\cap_{i=1}^{n-1} \mathfrak{A}_i) = R$.

Now suppose $b_1, \ldots, b_n \in R$. Our inductive hypothesis implies there exists a $b \in R$ such that $b \equiv b_i \bmod \mathfrak{A}_i$ for $i = 1, \ldots, n-1$. From the case $n = 2$, there exists an $a \in R$ such that $a \equiv b_n \bmod \mathfrak{A}_n$ and $a \equiv b \bmod \cap_{i=1}^{n-1} \mathfrak{A}_i$. But then $a \equiv b_i \bmod \mathfrak{A}_i$ for all $i = 1, \ldots, n$. ∎

Suppose $R = \mathbb{Z}$ and $z_1, \ldots, z_n$ are pairwise, relatively prime integers. Then $(z_i) + (z_j) = \mathbb{Z}$ for any $i \neq j$. For any integers $b_1, \ldots, b_n \in \mathbb{Z}$, Theorem 14.1 implies there is a $z \in \mathbb{Z}$ such that $z \equiv b_i \bmod (z_i)$ for every $i = 1, \ldots, n$. This is the familiar version of the Chinese Remainder Theorem from elementary number theory.

If two ideals $\mathfrak{A}$ and $\mathfrak{B}$ of R have the property that $\mathfrak{A} + \mathfrak{B} = R$, then $\mathfrak{A}$ and $\mathfrak{B}$ are said to be comaximal. A set of ideals $\{\mathfrak{A}_1, \ldots, \mathfrak{A}_n\}$ of R is said to be pairwise comaximal if $\mathfrak{A}_i + \mathfrak{A}_j = R$ whenever $i \neq j$. The Chinese Remainder Theorem says if $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ are pairwise comaximal, then $R/\cap_{i=1}^{n} \mathfrak{A}_i \cong R/\mathfrak{A}_1 \oplus \cdots \oplus R/\mathfrak{A}_n$.

Special PIRs will play an important role in the basic structure theorem.

Definition 14.3 A principal ideal ring R is said to be special if R contains precisely one prime ideal $\mathfrak{P}$.

Recall an ideal $\mathfrak{A}$ of R is said to be nilpotent if $\mathfrak{A}^r = (0)$ for some positive integer r. The unique prime $\mathfrak{P}$ in a special PIR is necessarily nilpotent by Theorem 6.6. There is a simple way to manufacture examples of special PIRs.

Example 14.4 Suppose D is any PID (e.g., one of the examples listed in 1.7). The reader will recall from elementary algebra that D is a unique factorization domain. Let p be a prime (i.e., irreducible) element of D. Set $\mathfrak{Q} = (p^r) = Dp^r$. $\mathfrak{Q}$ is a primary ideal of D belonging to $\mathfrak{P} = Dp$. The only prime ideal in $R = D/\mathfrak{Q}$ is $\mathfrak{P}/\mathfrak{Q}$, and $(\mathfrak{P}/\mathfrak{Q})^r = (0)$. Thus, $R = D/\mathfrak{Q}$ is a special PIR.

For a concrete example, $\mathbb{Z}/2^r\mathbb{Z}$ is a special PIR. ∎

Theorem 14.5 Let R be a principal ideal ring.

(a) If $\mathfrak{P}$ and $\mathfrak{P}'$ are prime ideals of R such that $\mathfrak{P}' < \mathfrak{P}$, then $\mathfrak{P}$ contains only two prime ideals $\mathfrak{P}$ and $\mathfrak{P}'$.

(b) If $\mathfrak{P}$ and $\mathfrak{P}'$ are prime ideals of R such that $\mathfrak{P}' < \mathfrak{P}$, then any primary ideal of R which is contained in $\mathfrak{P}$ contains $\mathfrak{P}'$.

(c) If $\mathfrak{P}'$ is a prime ideal of R which is not maximal, then the only primary ideal of R contained in $\mathfrak{P}'$ is $\mathfrak{P}'$ itself.

(d) Any two prime ideals of R are either comaximal or one contains the other.

Proof. We prove (b) first. Since R is a PIR, $\mathfrak{P} = Rp$, and $\mathfrak{P}' = Rp'$ for some elements p and p' in R. Since $\mathfrak{P}' \subseteq \mathfrak{P}$, $p' = pr$ for some $r \in R$. The element p is not in $\mathfrak{P}'$ since $\mathfrak{P}' \neq \mathfrak{P}$. Since $\mathfrak{P}'$ is a prime ideal of R, $r \in \mathfrak{P}' = Rp'$. Therefore, $r = sp'$ for some $s \in R$. In particular, $p' = rp = spp'$.

Now let $\mathfrak{Q}$ be any primary ideal of R such that $\mathfrak{Q} \subseteq \mathfrak{P}$. Then $(1 - sp)p' = 0 \in \mathfrak{Q}$. Since $1 - sp$ is $\notin \mathfrak{P}$, $1 - sp \notin \sqrt{\mathfrak{Q}}$. Since $\mathfrak{Q}$ is primary, $p' \in \mathfrak{Q}$. Therefore, $\mathfrak{P}' \subseteq \mathfrak{Q}$.

(a) Since $\mathfrak{P}'$ is a primary ideal contained in $\mathfrak{P}$, (b) implies $\mathfrak{P}'$ is the intersection of all primary ideals contained in $\mathfrak{P}$. In particular, $\mathfrak{P}'$ is uniquely determined by $\mathfrak{P}$. This proves (a).

(c) Suppose $\mathfrak{P}'$ is a prime ideal of R which is not maximal. Then there exists a prime ideal $\mathfrak{P}$ such that $\mathfrak{P}' < \mathfrak{P} < R$. Suppose $\mathfrak{Q}$ is a primary ideal of R contained in $\mathfrak{P}'$. The argument in the second paragraph of this proof implies $\mathfrak{P}' \subseteq \mathfrak{Q}$. Therefore, $\mathfrak{Q} = \mathfrak{P}'$.

(d) Suppose $\mathfrak{P}_1$ and $\mathfrak{P}_2$ are two distinct prime ideals of R. If $\mathfrak{P}_1 + \mathfrak{P}_2 \neq R$, then $\mathfrak{P}_1, \mathfrak{P}_2 \subseteq \mathfrak{P}_1 + \mathfrak{P}_2 \subseteq \mathfrak{P}$ for some prime ideal $\mathfrak{P}$ of R. By (a), $\mathfrak{P}$ contains at most two prime ideals one of which must be $\mathfrak{P}$. Hence, one of the $\mathfrak{P}_i$ is $\mathfrak{P}$, and the other is contained in $\mathfrak{P}$. ∎

We can now prove the following structure theorem.

Theorem 14.6 Any principal ideal ring is a finite direct sum of principal ideal domains and special principal ideal rings.

Proof. Let R be a PIR. Then R is Noetherian. In particular, (0) has an irredundant primary decomposition: $(0) = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$. Set $\mathfrak{P}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$. We first observe that $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ are pairwise comaximal (prime) ideals of R. To see this, suppose $\mathfrak{P}_i + \mathfrak{P}_j \neq R$ for some $i \neq j$. By Theorem 14.5d, $\mathfrak{P}_i \subseteq \mathfrak{P}_j$, or $\mathfrak{P}_j \subseteq \mathfrak{P}_i$. Let us suppose $\mathfrak{P}_i \subseteq \mathfrak{P}_j$. Since the associated primes of an irredundant decomposition are all distinct, we have $\mathfrak{P}_i <$

$\mathfrak{P}_j < R$. But then Theorem 14.5b implies $\mathfrak{Q}_i \subseteq \mathfrak{P}_i \subseteq \mathfrak{Q}_j$. This is impossible since $(0) = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ is an irredundant decomposition of (0). Thus, for each $i \neq j$, $\mathfrak{P}_i + \mathfrak{P}_j = R$.

Suppose $1 \leq i < j \leq n$. Since $\mathfrak{P}_i + \mathfrak{P}_j = R$, we have

$$\sqrt{\mathfrak{Q}_i + \mathfrak{Q}_j} = \sqrt{\sqrt{\mathfrak{Q}_i} + \sqrt{\mathfrak{Q}_j}} = \sqrt{\mathfrak{P}_i + \mathfrak{P}_j} = \sqrt{R} = R$$

In particular, $\mathfrak{Q}_i + \mathfrak{Q}_j = R$. Thus, the primary ideals $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_n$ are pairwise comaximal. The Chinese Remainder Theorem implies

14.7 $R \cong R/(0) \cong R/\mathfrak{Q}_1 \oplus \cdots \oplus R/\mathfrak{Q}_n$.

Each ring $R/\mathfrak{Q}_i$ is a homomorphic image of R. In particular, each $R/\mathfrak{Q}_i$ is a PIR. If $\mathfrak{P}_i$ is a maximal ideal of R, then $\sqrt{\mathfrak{Q}_i} = \mathfrak{P}_i$ implies $\mathfrak{Q}_i$ is contained in only one prime ideal, namely $\mathfrak{P}_i$. In particular, $R/\mathfrak{Q}_i$ has precisely one prime ideal $\mathfrak{P}_i/\mathfrak{Q}_i$. Since $\mathfrak{P}_i$ is principal, $\mathfrak{P}_i^r \subseteq \mathfrak{Q}_i$ for some $r > 0$. Therefore, $\mathfrak{P}_i/\mathfrak{Q}_i$ is nilpotent. We conclude that $R/\mathfrak{Q}_i$ is a special PIR.

Suppose $\mathfrak{P}_i$ is not a maximal ideal of R. Theorem 14.5c implies $\mathfrak{Q}_i = \mathfrak{P}_i$. In particular, $R/\mathfrak{Q}_i = R/\mathfrak{P}_i$ is a PID. This completes the proof of Theorem 14.16. ∎

EXERCISES

1. Suppose $R_1, \ldots, R_n$ are PIRs. Show $R_1 \oplus \cdots \oplus R_n$ is a PIR.
2. In Exercise 1, is $R_1 \otimes_{\mathbb{Z}} \cdots \otimes_{\mathbb{Z}} R_n$ a PID?
3. In the proof of Theorem 14.6, we used the following result: Let $\mathfrak{A}$ and $\mathfrak{B}$ be ideals of a commutative ring R. Then

   $$\sqrt{\sqrt{\mathfrak{A}} + \sqrt{\mathfrak{B}}} = \sqrt{\mathfrak{A} + \mathfrak{B}}.$$

   Give a proof of this assertion.
4. Suppose R is a special PIR with unique prime $\mathfrak{P} = Rp$. Suppose $r > 1$ is the index of nilpotency of $\mathfrak{P}$. Show $\{Rp^k \mid k = 0, \ldots, r\}$ is the complete set of ideals of R.
5. Suppose p is a nilpotent element of R. If every $x \in R$ can be written in the form $x = \varepsilon p^k$ for some unit ε and nonnegative integer k, show R is a special PIR.
6. Let R be a PIR. Suppose M is a finitely generated R-module with basis $\Gamma = \{m_1, \ldots, m_n\}$ Show that every submodule of M is generated as an R-module by n or fewer elements.
7. Suppose R is a PID. Show every submodule of R^n is a free R-module of rank $\leq n$. (We will prove this in the next chapter.)
8. Is Exercise 7 true if R is only a PIR?

9. Suppose R is a PID. Let $x_1, \ldots, x_n \in R$, and suppose $Rx_1 + \cdots + Rx_n = Rd$. Show there exists a matrix $A \in M_{n\times n}(R)$ such that $\det(A) = d$ and $\mathrm{Row}_1(A) = (x_1, \ldots, x_n)$.
10. Use the results in Exercise 9 to prove the following result of H. J. S. Smith: Suppose R is a PID. Let $A \in M_{n\times n}(R)$. Then there exist invertible matrices $P, Q \in \mathrm{Gl}(n, R)$ such that PAQ is diagonal. (We will prove a more general result in the next chapter.)

15
The Smith Normal Form of a Matrix

In this chapter R will denote a commutative ring.

Definition 15.1 A 1×2 matrix $(a,b) \in M_{1\times 2}(R)$ admits a diagonal reduction if there exists an invertible matrix $Q \in \mathrm{Gl}(2,\mathrm{R})$ such that $(a,b)Q = (d,0)$ for some $d \in R$.

Notice if (a,b) admits a diagonal reduction, then the two generated ideal $Ra + Rb$ in R is principal. For suppose

$$(a,b)\begin{bmatrix} x & z \\ y & w \end{bmatrix} = (d,0) \quad \text{with } Q = \begin{bmatrix} x & z \\ y & w \end{bmatrix} \in \mathrm{Gl}(2,R)$$

If

$$Q^{-1} = \begin{bmatrix} x' & z' \\ y' & w' \end{bmatrix}$$

then

$$(a,b) = (d,0)\begin{bmatrix} x' & z' \\ y' & w' \end{bmatrix}$$

implies $a = dx'$, $b = dz'$, and $ax + by = d$. Therefore, $Ra + Rb = Rd$.

A ring R for which every 1×2 matrix in $M_{1\times 2}(R)$ admits a diagonal reduction is called a Hermite ring. Our comments above imply every two generated ideal in a Hermite ring is in fact principal. It follows easily from this that every finitely generated ideal in a Hermite ring is principal. In particular, if R is a Noetherian, Hermite ring, then R is a PIR.

Our first theorem in this section is due to I. Kaplansky. It provides us with a partial converse to the statement above.

Theorem 15.2 Let R be a PIR. Suppose $Z(R) \subseteq J(R)$. Then every 1×2 matrix in $M_{1\times 2}(R)$ admits a diagonal reduction.

Proof. Let $(a,b) \in M_{1\times 2}(R)$. Since R is a PIR, $Ra + Rb = Rd$ for some $d \in R$. If $d = 0$, then $a = b = 0$, and there is nothing to prove. Hence, we can assume $d \neq 0$. We introduce the following notation:

15.3

$$\begin{array}{lll} d = ap + bq & Ra_1 \cap Rb_1 = Rc & Rr + Rs = Rt \\ a = da_1 & c = a_1 r & r = xt \\ b = db_1 & c = -b_1 s & s = yt \end{array}$$

Since R is a PIR, elements $p, q, a_1, b_1, c, r, s, t, x$, and y can be found in R such that the relations in equation 15.3 are satisfied. Since $d \neq 0$, and $0 = ap + bq - d = d(a_1 p + b_1 q - 1)$, $a_1 p + b_1 q - 1 \in Z(R)$. Thus, $a_1 p + b_1 q - 1 \in J(R)$ by hypothesis. In particular, $u = a_1 p + b_1 q = 1 + [a_1 p + b_1 q - 1]$ is a unit in R. Set $p' = u^{-1}p$ and $q' = u^{-1}q$. Then we have the following equations.

15.4 $u^{-1}d = ap' + bq'$ and $a_1 p' + b_1 q' = 1$.

Set $z = a_1 p' b_1 = b_1(1 - b_1 q')$. Then $z \in Ra_1 \cap Rb_1 = Rc$. Let $z = cw = -b_1 sw$. If $b_1 \in J(R)$, then $a_1 p' = 1 - b_1 q' \in U(R)$. Thus, $a_1 \in U(R)$. The equations in 15.3 then imply $a \sim d$, and $b = a\alpha$ for some $\alpha \in R$. Then

$$(a,b)\begin{bmatrix} 1 & -\alpha \\ 0 & 1 \end{bmatrix} = (a,a\alpha)\begin{bmatrix} 1 & -\alpha \\ 0 & 1 \end{bmatrix} = (a,0),$$

a diagonal matrix. Thus, the proof is complete if $b_1 \in J(R)$. Hence, we assume $b_1 \notin J(R)$.

Now

$$\begin{aligned} b_1(1 - b_1 q' + sw) &= b_1(1 - b_1 q') + b_1 sw = z + b_1 sw \\ &= -b_1 sw + b_1 sw = 0 \end{aligned}$$

Since $Z(R) \subseteq J(R)$ and $b_1 \notin J(R)$, b_1 is a regular element of R. Therefore, $1 - b_1q' + sw = 0$. If $s \in J(R)$, then $1 + sw \in U(R)$, and hence $b_1 \in U(R)$.But then the equations in 15.3 imply $b \sim d$, and $a = b\alpha$ for some $\alpha \in R$. Then

$$(a,b)\begin{bmatrix} 0 & 1 \\ 1 & -\alpha \end{bmatrix} = (b\alpha,b)\begin{bmatrix} 0 & 1 \\ 1 & -\alpha \end{bmatrix} = (b,0)$$

So again the theorem is proved. Hence, we can assume $s \notin J(R)$.

Since $s = yt$, $t \notin J(R)$. We also have $c = a_1xt = -b_1yt$. Since $Z(R) \subseteq J(R)$, t is a regular element of R. Therefore, $a_1x = -b_1y$. In particular, $a_1x \in Ra_1 \cap Rb_1 = Rc$. Thus, $a_1x = -b_1y = ch$ for some $h \in R$. But $c = a_1xt$. Therefore, $ch = a_1xth$. Thus, $a_1x = -b_1y = ch = a_1xth$. If $a_1x = 0$, then $c = a_1r = a_1xt = 0$. Then $0 = cw = z = b_1(1 - b_1q')$. Since $b_1 \notin J(R)$, b_1 is a regular element of R. The last equation implies $b_1 \in U(R)$. We have seen above that (a,b) admits a diagonal reduction when b_1 is a unit. Hence, the theorem is proved if $a_1x = 0$. Consequently, we may assume $a_1x \neq 0$.

We have seen that $a_1x = a_1xth$. Therefore, $0 = a_1x(1 - th)$. Since $a_1x \neq 0$, $1 - th \in Z(R) \subseteq J(R)$. We conclude $t \in U(R)$. The equations in 15.3 then imply $Rr + Rs = R$. In particular,

15.5 $er + fs = 1$ for some $e, f \in R$.

Set $g = ep' + fq'$. It then follows from equations 15.4 and 15.5 that

$$\begin{bmatrix} a_1 & b_1 \\ e - ga_1 & f - gb_1 \end{bmatrix}\begin{bmatrix} p' & r \\ q' & s \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Thus,

$$\begin{bmatrix} p' & r \\ q' & s \end{bmatrix} \in \mathrm{Gl}(2,R)$$

Also,

$$(a,b)\begin{bmatrix} p' & r \\ q' & s \end{bmatrix} = (u^{-1}d,0)$$

Hence, (a,b) admits a diagonal reduction. This completes the proof of Theorem 15.2. ∎

The argument given in the proof of Theorem 15.2 works even for noncommutative rings. If T is a (not necessarily commutative) ring for which (1) $Z(T) \subseteq J(T)$, (2) the sum and intersection of any two principal right ideals is a principal right ideal of T, (3) the sum of any two principal left ideals is a principal left ideal of T, and (4) any matrix in $M_{2\times 2}(T)$ with a one-sided inverse is invertible, then T is a right Hermite ring; that is, for all $(a,b) \in M_{1\times 2}(T)$, there exists a $Q \in \mathrm{Gl}(2,T)$ such that $(a,b)Q = (d,0)$ for some $d \in T$. Only minor adjustments in the proof of Theorem 15.2 are needed to prove this result for noncommutative rings.

If every 1×2 matrix in $M_{1\times 2}(R)$ admits a diagonal reduction, then every 2×1 matrix in $M_{2\times 1}(R)$ also admits a diagonal reduction. By this, we mean for every

$$\begin{bmatrix} a \\ b \end{bmatrix} \in M_{2\times 1}(R)$$

there exists a $P \in \mathrm{Gl}(2,R)$ such that

$$P\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} d \\ 0 \end{bmatrix} \quad \text{for some } d \in R$$

This assertion follows easily from taking transposes. Thus, if R is a PIR such that $Z(R) \subseteq J(R)$, then every 1×1, 1×2, and 2×1 matrix with entries from R admits a diagonal reduction.

If R is a PID, then $Z(R) = (0)$. In particular, $Z(R) \subseteq J(R)$. Therefore, Theorem 15.2 implies any PID is a Hermite ring. If R is a special PIR, then R has only one prime ideal $\mathfrak{P}$ which is necessarily nilpotent. Therefore, $Z(R) \subseteq \mathfrak{P} = J(R)$. Thus, any special, PIR is a Hermite ring by Theorem 15.2. Theorem 14.6 implies any PIR is a finite direct sum of Hermite rings.

Before stating the main results of this section, we review some familiar definitions. Recall (Definition 2.23) that an $m \times n$ matrix D is a diagonal matrix if $[D]_{ij} = 0$ whenever $i \neq j$. If D is a diagonal matrix in $M_{m\times n}(R)$, then we will write D in the following way: $D = \mathrm{Diag}(d_1, \ldots, d_r)$. Here $r = \min\{m,n\}$, and $d_i = [D]_{ii}$ for $i = 1, \ldots, r$. Notice that a diagonal matrix need not be square. If $m \leq n$, then

$$\mathrm{Diag}(d_1, \ldots, d_r) = \begin{bmatrix} d_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & d_r & \cdots & 0 \end{bmatrix}$$

If $m \geq n$, then

$$\text{Diag}(d_1, \ldots, d_r) = \begin{bmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & d_r \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}$$

Definition 15.6 Let $A, B \in M_{m\times n}(R)$. The matrices A and B are said to be equivalent if $PAQ = B$ for some $P \in \text{Gl}(m,R)$ and $Q \in \text{Gl}(n,R)$.

If A and B are equivalent, we will write $A \approx B$. It is easy to check that $\approx$ is an equivalence relation on $M_{m\times n}(R)$. Thus, for all A, B, and $C \in M_{m\times n}(R)$, $A \approx A$, $A \approx B$ if and only if $B \approx A$, and if $A \approx B$ and $B \approx C$ then $A \approx C$. Notice that only matrices of the same size can be equivalent.

Definition 15.7 A commutative ring R is called an elementary divisor ring if for all $m,n \geq 1$ and for every $A \in M_{m\times n}(R)$, there exists a diagonal matrix $\text{Diag}(d_1, \ldots, d_r) \in M_{m\times n}(R)$ such that

(a) $A \approx \text{Diag}(d_1, \ldots, d_r)$, and

(b) $d_i \mid d_{i+1}$ for all $i = 1, \ldots, r-1$. [Here $r = \min\{m,n\}$.]

The reader will recall that the notation $d_i \mid d_{i+1}$ in 15.7b means d_i divides d_{i+1}. Thus, $d_i z_i = d_{i+1}$ for some $z_i \in R$.

A matrix $A \in M_{m\times n}(R)$ is said to have a diagonal reduction if there exist invertible matrices $P \in \text{Gl}(m,R)$ and $Q \in \text{Gl}(n,R)$ such that $PAQ = \text{Diag}(d_1, \ldots, d_r)$ $[r = \min\{m,n\}]$, with $d_i \mid d_{i+1}$ for all $i = 1, \ldots, r-1$. Thus, R is an elementary divisor ring if for all $m,n \geq 1$, every $m \times n$ matrix in $M_{m\times n}(R)$ has a diagonal reduction. If $D = \text{Diag}(d_1, \ldots, d_r) \in M_{m\times n}(R)$, and $d_i \mid d_{i+1}$ for all $i = 1, \ldots, r-1$, then D is called a diagonal reduction of A if $A \approx D$.

We will abbreviate the expression $d_i \mid d_{i+1}$ for all $i = 1, \ldots, r-1$ by writing $d_1 \mid d_2 \mid \cdots \mid d_r$. Thus, R is an elementary divisor ring if every $m \times n$ matrix A with entries in R is equivalent to some diagonal matrix $\text{Diag}(d_1, \ldots, d_r)$ in which $d_1 \mid d_2 \mid \cdots \mid d_r$.

Suppose $A \in M_{m\times n}(R)$ has a diagonal reduction $D = \text{Diag}(d_1, \ldots, d_r)$. Then $A \approx D$ in $M_{m\times n}(R)$, and $d_1 \mid d_2 \mid \cdots \mid d_r$ in R. Notice that $d_1 \mid d_2 \mid \cdots \mid d_r$ in R if and only if $Rd_1 \supseteq Rd_2 \supseteq \cdots \supseteq Rd_r$. In particular, if d_i is a unit in R, then $d_1, \ldots, d_i$ are all units in R. Similarly, if $d_i = 0$, then $d_i = d_{i+1} = \cdots = d_r = 0$. Thus, if $D = \text{Diag}(d_1, \ldots, d_r)$ is a diagonal reduction of A, then the diagonal entries $d_1, \ldots, d_r$ of D are arranged as follows: Those d_i (if any) which are units in R appear in the upper left-hand corner of the diagonal of D. Those d_i (if any) which are nonzero, nonunits of R appear next on the diagonal of D. Those d_i (if any) which are zero appear in the lower right-hand portion of the diagonal of D.

We certainly know one familiar example of an elementary divisor ring. Any field F is an elementary divisor ring. To see this, suppose $A \in M_{m\times n}(F)$. Then A can be reduced by a finite number of elementary row and column operations to an $m \times n$ matrix of the form

$$\left[\begin{array}{c|c} I_k & \mathrm{O} \\ \hline \mathrm{O} & \mathrm{O} \end{array}\right]$$

Here k is the rank of A. Since elementary row and column operations are performed on A by multiplying A on the left and right by suitable invertible matrices, there exist $P \in \text{Gl}(m,F)$ and $Q \in \text{Gl}(n,F)$ such that

$$PAQ = \left[\begin{array}{c|c} I_k & \mathrm{O} \\ \hline \mathrm{O} & \mathrm{O} \end{array}\right]$$

Thus, any field is an elementary divisor ring.

We will show that any PIR is an elementary divisor ring. This follows from our second theorem, which is also due to I. Kaplansky.

Theorem 15.8 Any Noetherian, Hermite ring is an elementary divisor ring.

Proof. Suppose R is a Noetherian, Hermite ring. Since R is Hermite, all 1×1, 1×2, and 2×1 matrices with entries from R admit diagonal reductions. Hence, we first argue any 2×2 matrix with entries from R admits a diagonal reduction.

Let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in M_{2\times 2}(R)$$

Set $E(A) = \{PAQ \mid P,Q \in \text{Gl}(2,R)\}$. The set $E(A)$ is just the equivalence class of A in $(M_{2\times 2}(R),\approx)$. For each $B \in E(A)$, consider the principal ideal $([B]_{11}) = R[B]_{11}$ in R. Let $\mathcal{S} = \{R[B]_{11} \mid B \in E(A)\}$. $\mathcal{S}$ is a nonempty set of ideals

of R. Since R is a Noetherian, $\mathcal{S}$ has a maximal element with respect to inclusion (Lemma 10.2). Hence, there exists a

$$W = \begin{bmatrix} d_1 & x \\ y & z \end{bmatrix} \in E(A)$$

such that Rd_1 is not strictly contained in $R[B]_{11}$ for any $B \in E(A)$. We claim $d_1 \mid x$ and $d_1 \mid y$.

Since R is Hermite, there exists a $Q \in \mathrm{Gl}(2,R)$ such that $(d_1,x)Q = (\alpha,0)$ for some $\alpha \in R$. We have noted previously that $R\alpha = Rd_1 + Rx$. Since

$$W \approx WQ = \begin{bmatrix} \alpha & 0 \\ * & * \end{bmatrix}$$

$R\alpha \in \mathcal{S}$. Since $Rd_1 \subseteq R\alpha$, the maximality of Rd_1 in $\mathcal{S}$ implies $Rd_1 = R\alpha$. Therefore, $d_1 \mid \alpha \mid x$. Similarly, there exists a $P \in \mathrm{Gl}(2,R)$ such that

$$P\begin{bmatrix} d_1 \\ y \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \end{bmatrix}$$

for some $\beta \in R$. Then $R\beta = Rd_1 + Ry$, and

$$W \approx PW = \begin{bmatrix} \beta & * \\ 0 & * \end{bmatrix}$$

The same argument as before shows $d_1 \mid \beta \mid y$.

Since $d_1 \mid x$ and $d_1 \mid y$, $W \approx \mathrm{Diag}(d_1,d_2)$ for some $d_2 \in R$. The P and Q used to reduce W to $\mathrm{Diag}(d_1,d_2)$ here are suitably chosen elementary matrices. Thus, $\mathrm{Diag}(d_1,d_2) \in E(A)$. Since

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} = \begin{bmatrix} d_1 & d_2 \\ 0 & d_2 \end{bmatrix}$$

we have

$$\begin{bmatrix} d_1 & d_2 \\ 0 & d_2 \end{bmatrix} \in E(A)$$

The same argument as above then shows $d_1 \mid d_2$. Thus, $A \approx W \approx \mathrm{Diag}(d_1,d_2)$ with $d_1 \mid d_2$.

We have now shown that a Noetherian, Hermite ring R has the property that all 1×1, 1×2, 2×1, and 2×2 matrices with entries from R have diagonal reductions. This will be enough to prove the theorem. In fact we have the following more general result:

Claim. For any commutative ring R, if all 1×1, 1×2, 2×1, and 2×2 matrices with entries from R have diagonal reductions, then all matrices with entries from R have diagonal reductions.

To see that this claim is true, let $A \in M_{m \times n}(R)$. By replacing A with A^t if need be, we can assume with no lost of generality that $m \geq n$. Proceeding by induction, we can assume the claim is true for smaller m, and for the given m, smaller n. We can also assume $m \geq 3$. Partition A as follows:

$$A = \left[\frac{A_1}{A_2}\right]$$

Here $A_1 \in M_{1 \times n}(R)$, and $A_2 \in M_{(m-1) \times n}(R)$. By our induction hypothesis, there exist $P_1 \in \mathrm{Gl}(m-1, R)$ and $Q_1 \in \mathrm{Gl}(n,R)$ such that $B = P_1 A_2 Q_1 = \mathrm{Diag}(x_1, \ldots, x_r)$ with $x_1 \mid x_2 \mid \cdots \mid x_r$. Here $r = \min\{m-1, n\}$. Then

$$\left[\begin{array}{c|c} 1 & \mathrm{O} \\ \hline \mathrm{O} & P_1 \end{array}\right] \left[\frac{A_1}{A_2}\right] Q_1 = \left[\frac{A_1 Q_1}{B}\right] = C$$

Since

$$\left[\begin{array}{c|c} 1 & \mathrm{O} \\ \hline \mathrm{O} & P_1 \end{array}\right]$$

is invertible, $A \approx C$. Partition

$$C = \left[\frac{D}{E}\right]$$

with $D \in M_{2 \times n}(R)$ and $E \in M_{(m-2) \times n}(R)$.

Again by our induction hypotheses there exist $P_2 \in \mathrm{Gl}(2,R)$ and $Q_2 \in \mathrm{Gl}(n,R)$ such that

$$F = P_2 D Q_2 = \begin{bmatrix} y_1 & 0 & \cdots & 0 \\ 0 & y_2 & \cdots & 0 \end{bmatrix}$$

with $y_1 \mid y_2$. Set

$$H = \left[\begin{array}{c|c} P_2 & \mathrm{O} \\ \hline \mathrm{O} & I_{m-2} \end{array}\right] \left[\frac{D}{E}\right] Q_2 = \left[\frac{F}{G}\right] = \left[\begin{array}{cccc} y_1 & 0 & \cdots & 0 \\ 0 & y_2 & \cdots & 0 \\ \hline \multicolumn{4}{c}{G} \end{array}\right]$$

Then $A \approx H$. Since y_1 divides all the entries in F, y_1 divides all the entries in $D = P_2^{-1}FQ_2^{-1}$. In particular, $y_1 \mid x_1$. Since $G = EQ_2$ and x_1 divides all the entries in B, x_1 divides all the entries in E. Therefore, x_1 divides all the entries in G. Since $y_1 \mid x_1$, y_1 divides all the entries in G. Thus, y_1 divides all the entries in H. Elementary row operations applied to H give us

$$H \approx \left[\begin{array}{c|c} y_1 & O \\ \hline O & K \end{array}\right]$$

with $K \in M_{(m-1)\times(n-1)}(R)$ and y_1 dividing all the entries of K. By our induction hypothesis, $K \approx \text{Diag}(z_1, \ldots, z_s)$ with $z_1 \mid z_2 \mid \cdots \mid z_s$ and $s = \min\{m - 1, n - 1\}$. If $P'KQ' = \text{Diag}(z_1, \ldots, z_s)$, then

$$\left[\begin{array}{c|c} 1 & O \\ \hline O & P' \end{array}\right]\left[\begin{array}{c|c} y_1 & O \\ \hline O & K \end{array}\right]\left[\begin{array}{c|c} 1 & O \\ \hline O & Q' \end{array}\right] = \text{Diag}(y_1, z_1, \ldots, z_s)$$

Thus, $A \approx H \approx \text{Diag}(y_1, z_1, \ldots, z_s)$ and $y_1 \mid z_1 \mid z_2 \mid \cdots \mid z_s$. This completes the proof of the claim and, consequently, the proof of the theorem. ∎

We can now combine Theorems 15.2 and 15.8 for our main result.

Theorem 15.9 Any principal ideal ring is an elementary divisor ring.

Proof. Let R be a PIR. Theorem 14.6 implies $R = R_1 \oplus \cdots \oplus R_p$ where each R_i is either a PID or a special PIR. In either case, $Z(R_i) \subseteq J(R_i)$ and, consequently, Theorem 15.2 implies R_i is a Hermite ring. Since each R_i is a PIR, each R_i is Noetherian. Thus, Theorem 15.8 implies each R_i is an elementary divisor ring. In particular, we need only argue that a finite direct sum of elementary divisor rings is again an elementary divisor ring. Since $R_iR_j = (0)$ whenever $i \neq j$, this is easy.

Fix $m,n \geq 1$, and let $A \in M_{m\times n}(R)$. Since $R = R_1 \oplus \cdots \oplus R_p$, $A = A_1 + \cdots + A_p$ with each $A_i \in M_{m\times n}(R_i)$ for $i = 1, \ldots, p$. Since each R_i is an elementary divisor ring, there exist $P_i \in \text{Gl}(m,R_i)$ and $Q_i \in \text{Gl}(n,R_i)$ such that $P_iA_iQ_i = \text{Diag}(d_{1i}, \ldots, d_{ri})$ with $d_{1i} \mid d_{2i} \mid \cdots \mid d_{ri}$ in R_i for $i = 1, \ldots, p$. Here $r = \min\{m,n\}$. Set $P = P_1 + \cdots + P_p$ and $Q = Q_1 + \cdots + Q_p$. Since $R_iR_j = (0)$ whenever $i \neq j$, it is easy to check that P and Q are invertible with inverses $P^{-1} = P_1^{-1} + \cdots + P_p^{-1}$ and $Q^{-1} = Q_1^{-1} + \cdots + Q_p^{-1}$. Thus, $P \in \text{Gl}(m,R)$ and $Q \in \text{Gl}(n,R)$. Finally, $PAQ = \text{Diag}(\sum_{i=1}^p d_{1i}, \ldots, \sum_{i=1}^p d_{ri})$, and $\sum_{i=1}^p d_{1i} \mid \sum_{i=1}^p d_{2i} \mid \cdots \mid \sum_{i=1}^p d_{ri}$. Thus, A has a diagonal reduction. We conclude that $R = R_1 \oplus \cdots \oplus R_p$ is an elementary divisor ring. ∎

Theorem 15.9 says that any matrix $A \in M_{m\times n}(R)$ has a diagonal reduction when R is a PIR. Such reductions are called Smith normal forms. To be more precise, we have the following definition.

Definition 15.10 Let $A \in M_{m\times n}(R)$. A diagonal matrix $D = \mathrm{Diag}(d_1, \ldots, d_r) \in M_{m\times n}(R)$ is called a Smith normal form of A if $A \approx D$ and $d_1 \mid d_2 \mid \cdots \mid d_r$ in R.

Thus, any diagonal reduction of A is called a Smith normal form of A. Theorem 15.9 guarantees that every $m \times n$ matrix with entries from a PIR has a Smith normal form.

The next order of business is to consider the question of uniqueness of the Smith normal form. Consider the following example.

Example 15.11 Let $R = \mathbb{Z}/6\mathbb{Z}$. Then R is a PIR. In particular, every $m \times n$ matrix with entries from R has a Smith normal form. Let us write the elements of R as $\{0,1,2,3,4,5\}$.

Let

$$A = \begin{bmatrix} 1 & 0 \\ 4 & 2 \end{bmatrix} \in M_{2\times 2}(R)$$

Set

$$P = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$$

Then

$$P^{-1} = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}$$

and $P^{-1}AP = \mathrm{Diag}(1,2)$. Thus,

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

is a Smith normal form of A.

On the other hand, if

$$Q = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$$

then

$$Q^{-1} = \begin{bmatrix} 2 & 3 \\ 5 & 2 \end{bmatrix}$$

and $Q^{-1}AQ = \text{Diag}(4,5)$. Then

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 4 \end{bmatrix}$$

Since $5 \in U(R)$,

$$\begin{bmatrix} 5 & 0 \\ 0 & 4 \end{bmatrix}$$

is another Smith normal form of A. Thus, there are at least two Smith normal forms of A in $M_{2\times 2}(R)$. ∎

We have seen in Example 15.11 that a given matrix may have more than one Smith normal form. Thus, in general, a Smith normal form of a matrix (if it exists) may not be unique. However, notice in Example 15.11 that $1 \sim 5$ and $2 \sim 4$ ($5 \times 2 \equiv 4 \bmod 6$). When R is a PIR, this is always the case. We will show that if $D = \text{Diag}(d_1, \ldots, d_r)$ and $D' = \text{Diag}(d_1, \ldots, d_r)$ are two Smith normal forms in $M_{m\times n}(R)$ and R is a PIR, then $D \approx D'$ implies $d_1 \sim d_1', \ldots, d_r \sim d_r'$. Thus, a Smith normal form of a matrix is unique up to associates along its diagonal when R is a PIR.

In order to prove the statements in the last paragraph, we need to discuss some well-known facts about cyclic modules. For this discussion, R can be any commutative ring. If M is a finitely generated R-module, we will let $\mu_R(M)$ denote the smallest number of elements in M which generate M as an R-module. For example, if R is a field, then $\mu_R(M)$ is just the vector space dimension of M over R. If R is a local ring, (R,m,k), then Nakayama's lemma (Corollary 10.20) implies $\mu_R(M) = \dim_k(M/mM)$. If $M = (0)$, then we set $\mu_R(0) = 0$.

Lemma 15.12 Suppose $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ are ideals in R such that $\mathfrak{A}_1 + \cdots + \mathfrak{A}_n \neq R$. Then $\mu_R(R/\mathfrak{A}_1 \oplus \cdots \oplus R/\mathfrak{A}_n) = n$.

Proof. Set $M = R/\mathfrak{A}_1 \oplus \cdots \oplus R/\mathfrak{A}_n$. Since $\mathfrak{A}_1 + \cdots + \mathfrak{A}_n \neq R$, there exists a maximal ideal m in R such that $\mathfrak{A}_1 + \cdots + \mathfrak{A}_n \subseteq m$. Clearly, $mM = m/\mathfrak{A}_1 \oplus \cdots \oplus m/\mathfrak{A}_n$ and $M/mM \cong R/m \oplus \cdots \oplus R/m$ (n summands). In particular, $\mu_R(M) \geq \dim_{R/m}(M/mM) = n$.

On the other hand, let $\delta_i = (0, \ldots, 1 + \mathfrak{A}_i, \ldots, 0) \in M$ for each $i = 1, \ldots, n$. Clearly, $\{\delta_1, \ldots, \delta_n\}$ is an R-module basis of M. Therefore, $\mu_R(M) \leq n$. ∎

Lemma 15.13 Suppose $\mathfrak{A}_1 \supseteq \mathfrak{A}_2 \supseteq \cdots \supseteq \mathfrak{A}_n$ and $\mathfrak{B}_1 \supseteq \mathfrak{B}_2 \supseteq \cdots \supseteq \mathfrak{B}_m$ are two sequences of ideals in R. We assume $\mathfrak{A}_1 \neq R \neq \mathfrak{B}_1$. If the R-modules $R/\mathfrak{A}_1$

$\oplus \cdots \oplus R/\mathfrak{A}_n$ and $R/\mathfrak{B}_1 \oplus \cdots \oplus R/\mathfrak{B}_m$ are isomorphic as R-modules, then $n = m$ and $\mathfrak{A}_i = \mathfrak{B}_i$ for all $i = 1, \ldots, n$.

Proof. Since $\bigoplus_{i=1}^n R/\mathfrak{A}_i \cong \bigoplus_{j=1}^m R/\mathfrak{B}_j$, Lemma 15.12 implies

$$n = \mu_R\left(\bigoplus_{i=1}^{n} R/\mathfrak{A}_i\right) = \mu_R\left(\bigoplus_{j=1}^{m} R/\mathfrak{B}_j\right) = m$$

Suppose some $\mathfrak{A}_j \neq \mathfrak{B}_j$ for some $j \in \{1, \ldots, n\}$. Let i be the smallest such j. Then $\mathfrak{A}_1 = \mathfrak{B}_1, \ldots, \mathfrak{A}_{i-1} = \mathfrak{B}_{i-1}$, and $\mathfrak{A}_i \neq \mathfrak{B}_i$. Here $1 \leq i \leq n$. We may assume that $\mathfrak{A}_i$ is not contained in $\mathfrak{B}_i$.

Let $x \in \mathfrak{A}_i - \mathfrak{B}_i$, and set $M = R/\mathfrak{A}_1 \oplus \cdots \oplus R/\mathfrak{A}_n$. For any $j = 1, \ldots, n$, $x(R/\mathfrak{A}_j) \cong R/(\mathfrak{A}_j : Rx)$. The isomorphism here is given by the map $1 + (\mathfrak{A}_j : Rx) \mapsto x + \mathfrak{A}_j$. Since $x \in \mathfrak{A}_i \subseteq \mathfrak{A}_{i-1} \subseteq \cdots \subseteq \mathfrak{A}_1$, $\mathfrak{A}_j : Rx = R$ for $j = 1, \ldots, i$. Therefore,

$$\textbf{15.14} \quad xM \cong \bigoplus_{j=i+1}^{n} R/(\mathfrak{A}_j : Rx)$$

Thus, xM is isomorphic to the direct sum of at most $n - i$ nonzero, cyclic R-modules. Since $\mathfrak{A}_{i+1} \supseteq \mathfrak{A}_{i+2} \supseteq \cdots \supseteq \mathfrak{A}_n$,

$$\mathfrak{A}_{i+1} : Rx \supseteq \mathfrak{A}_{i+2} : Rx \supseteq \cdots \supseteq \mathfrak{A}_n : Rx$$

Thus, Lemma 15.12 implies $\mu_R(xM) \leq n - i$.

We also have

$$\textbf{15.15} \quad xM \cong \bigoplus_{j=1}^{n} R/(\mathfrak{B}_j : Rx)$$

For any ideal $\mathfrak{A}$ in R, $\mathfrak{A} : Rx = R$ if and only if $x \in \mathfrak{A}$. Since $x \notin \mathfrak{B}_i$ and $\mathfrak{B}_i \supseteq \mathfrak{B}_{i+1} \supseteq \cdots \supseteq \mathfrak{B}_n$, the submodules $R/(\mathfrak{B}_i : Rx), \ldots, R/(\mathfrak{B}_n : Rx)$ are all nonzero. Thus, equation 15.15 implies $\mu_R(xM) \geq n - i + 1$. This is clearly impossible. We conclude that $\mathfrak{A}_j = \mathfrak{B}_j$ for all $j = 1, \ldots, n$. ∎

We can now use these two lemmas to address the uniqueness question for the Smith normal form.

Lemma 15.16 Let $D_1 = \text{Diag}(d_1, \ldots, d_r) \in M_{m\times n}(R)$ with $d_1 \mid d_2 \mid \cdots \mid d_r$. (As usual, $r = \min\{m,n\}$.) Let $D_2 = \text{Diag}(s_1, \ldots, s_r) \in M_{m\times n}(R)$ with $s_1 \mid s_2 \mid \cdots \mid s_r$. If $D_1 \approx D_2$, then $Rd_i = Rs_i$ for all $i = 1, \ldots, r$.

Proof. Let $M_1 = R/Rd_1 \oplus \cdots \oplus R/Rd_r \oplus R^{n-r}$. Here $R^{n-r} = R \oplus \cdots \oplus R$ ($n - r$ times). If $r = n$ (i.e., $n \leq m$), then $M_1 = R/Rd_1 \oplus \cdots \oplus R/Rd_r$. Some of the summands R/Rd_i could very well be zero here. Clearly, $R/Rd_i = (0)$

if and only if $d_i \in U(R)$. Even if $R/Rd_i = (0)$, we will keep this summand present in our notation in what follows.

We have the following short exact sequence for M_1.

$$\textbf{15.17}\quad (0) \mapsto K \overset{\iota}{\mapsto} R^n \overset{f_1}{\mapsto} M_1 \mapsto (0)$$

The map f_1 is the R-module homomorphism given by

$$f_1(\varepsilon_i) = \begin{cases} (0, \ldots, 0, \underset{i}{1 + Rd_i}, 0, \ldots, 0) & \text{for } i = 1, \ldots, r \\ (0, \ldots, 0, \underset{r}{0}, 0, \ldots, 0, \underset{i}{1}, 0, \ldots, 0) & \text{for } i = r+1, \ldots, n \end{cases}$$

In this formula, the r or i below an entry of the n-tuple indicates what summand contains the entry. As usual, $\varepsilon = \{\varepsilon_1, \ldots, \varepsilon_n\}$ is the canonical basis of R^n. $K = \operatorname{Ker}(f_1)$ and ι is the inclusion map. Clearly, $K = R(d_1\varepsilon_1) \oplus \cdots \oplus R(d_r\varepsilon_r)$.

The matrix $D_1 = \operatorname{Diag}(d_1, \ldots, d_r)$ induces an R-module homomorphism $g_1 : M_{1\times m}(R) \mapsto K$ given by $g_1(x_1, \ldots, x_m) = ((x_1, \ldots, x_m)D_1)^t$. To see this, first observe that the map $g_1(\xi) = (\xi D_1)^t = D_1^t\xi^t$ is certainly an R-module homomorphism from $M_{1\times m}(R)$ to R^n. For $j = 1, \ldots, m$, let $\delta_j = (0, \ldots, 1, \ldots, 0)$ be the row vector in $M_{1\times m}(R)$ having a 1 in its jth entry. Then we have

$$\textbf{15.18}\quad g_1(\delta_j) = (\delta_j D_1)^t = \begin{cases} d_j\varepsilon_j & \text{for } j = 1, \ldots, r \\ 0 & \text{for } j = r+1, \ldots, m \end{cases}$$

Therefore, $\operatorname{Im}(g_1) = K$. Since $M_{1\times m}(R)$ is a free R-module of rank m,

$$\textbf{15.19}\quad M_{1\times m}(R) \overset{g_1}{\mapsto} R^n \overset{f_1}{\mapsto} M_1 \mapsto (0)$$

is a presentation of the R-module M_1.

Set $M_2 = R/Rs_1 \oplus \cdots \oplus R/Rs_r \oplus R^{n-r}$. Then we can find a similar presentation for M_2.

$$\textbf{15.20}\quad M_{1\times m}(R) \overset{g_2}{\mapsto} R^n \overset{f_2}{\mapsto} M_2 \mapsto (0)$$

The R-module homomorphisms in 15.20 are as follows:

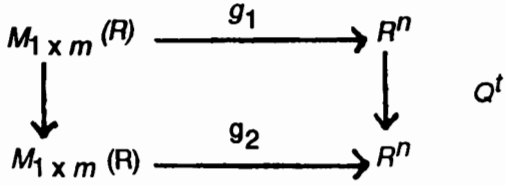
$$g_2(\xi) = (\xi D_2)^t \quad \text{for all } \xi \in M_{1\times m}(R)$$

and

$$f_2(\varepsilon_i) = \begin{cases} (0, \ldots, 0, \underset{i}{1 + Rs_i}, 0, \ldots, 0) & \text{for } i = 1, \ldots, r \\ (0, \ldots, 0, \underset{r}{0}, 0, \ldots, 0, \underset{i}{1}, 0, \ldots, 0) & \text{for } i = r+1, \ldots, n \end{cases}$$

Since $D_1 \approx D_2$, there exist invertible matrices $P \in \mathrm{Gl}(m,R)$ and $Q \in \mathrm{Gl}(n,R)$ such that $PD_1Q = D_2$. We claim the following diagram is commutative:

15.21



The vertical map Q^t in 15.21 means left multiplication by Q^t. The vertical map P^{-1} means right multiplication by P^{-1}. Since these matrices are invertible, both of the vertical arrows in diagram 15.21 are R-module isomorphisms.

Let $\xi \in M_{1\times m}(R)$. Then $Q^t g_1(\xi) = Q^t D_1^t \xi^t$. On the other hand, $g_2(\xi P^{-1}) = [(\xi P^{-1})D_2]^t = D_2^t (P^{-1})^t \xi^t$. Since $PD_1Q = D_2$, $D_1Q = P^{-1}D_2$ and $Q^tD_1^t = D_2^t(P^{-1})^t$. Thus, $Q^t g_1(\xi) = g_2(\xi P^{-1})$. In particular, the diagram in 15.21 is commutative.

Since the diagram in 15.21 is commutative, we have

$$\begin{aligned} Q^t(\mathrm{Im}(g_1)) &= Q^t(g_1(M_{1\times m}(R))) = Q^t g_1(M_{1\times m}(R)) \\ &= g_2 P^{-1}(M_{1\times m}(R)) = g_2(M_{1\times m}(R)P^{-1}) \\ &= g_2(M_{1\times m}(R)) = \mathrm{Im}(g_2) \end{aligned}$$

Thus, the isomorphism $Q^t : R^n \mapsto R^n$ maps $\mathrm{Im}(g_1)$ onto $\mathrm{Im}(g_2)$. The isomorphism theorems imply $R^n/\mathrm{Im}(g_1) \cong R^n/\mathrm{Im}(g_2)$. We now have the following sequence of R-module isomorphisms:

$$M_1 \cong R^n/\mathrm{Ker}(f_1) = R^n/\mathrm{Im}(g_1) \cong R^n/\mathrm{Im}(g_2) = R^n/\mathrm{Ker}(f_2) \cong M_2$$

Thus, the R-modules M_1 and M_2 are isomorphic.

We can now apply Lemma 15.13. Since $d_1 \mid d_2 \mid \cdots \mid d_r$, we have $Rd_1 \supseteq Rd_2 \supseteq \cdots \supseteq Rd_r$. Thus, the units (if any) in the sequence $\{d_1, \ldots, d_r\}$ are listed first. A similar statement is true for the sequence $\{s_1, \ldots, s_r\}$. Let l_1 and l_2 denote the numbers of units in $\{d_1, \ldots, d_r\}$ and $\{s_1, \ldots, s_r\}$, respectively. Then $0 \le l_i \le r$ for each $i = 1,2$.

Suppose $l_1 = r$, that is, $d_1, \ldots, d_r \in U(R)$. Then $M_1 = R/Rd_1 \oplus \cdots \oplus R/Rd_r \oplus R^{n-r} = R^{n-r}$ (or (0) if $r = n$). Since $M_2 \cong M_1$, we have $R/Rs_1 \oplus \cdots \oplus R/Rs_r \oplus R^{n-r} \cong R^{n-r}$ (or (0) if $r = n$). It follows easily from Lemma 15.12 that $R/Rs_i = (0)$ for all $i = 1, \ldots, r$. Thus, $s_1, \ldots, s_r \in U(R)$, and, in particular, $Rd_i = R = Rs_i$ for all $i = 1, \ldots, r$. Hence, if $l_1 = r$, the proof of the lemma is complete.

Suppose $l_1 < r$. By reversing the roles of D_1 and D_2, we also have $l_2 < r$. Then

15.22 $$M_1 = R/Rd_{l_1+1} \oplus \cdots \oplus R/Rd_r \oplus R^{n-r}$$
$$M_2 = R/Rs_{l_2+1} \oplus \cdots \oplus R/Rs_r \oplus R^{n-r}$$

These two direct sum decompositions correspond to the following descending chains of ideals in R.

15.23 $$R > Rd_{l_1+1} \supseteq \cdots \supseteq Rd_r \supseteq (0) \supseteq \cdots \supseteq (0)$$
$$R > Rs_{l_2+1} \supseteq \cdots \supseteq Rs_r \supseteq (0) \supseteq \cdots \supseteq (0)$$

The zeros in 15.23 appear $n - r$ times if $r < n$. If $r = n$, then these sequences end with Rd_r and Rs_r, respectively. (Of course, some of the Rd_i or Rs_i could be zero as well.) At any rate, since $M_1 \cong M_2$, Lemma 15.13 implies

$$\begin{aligned}(r - l_1) + (n - r) &= \mu_R(M_1) = \mu_R(M_2) \\ &= (r - l_2) + (n - r)\end{aligned}$$

Therefore, $l_1 = l_2$. We also get $Rd_i = Rs_i$ for all $i = l_1 + 1, \ldots, r$. In particular, $Rd_i = Rs_i$ for all $i = 1, \ldots, r$. This completes the proof of the lemma. ∎

We can now apply Lemma 15.16 to the case when R is a PIR.

Theorem 15.24 Let R be a PIR. Then any $A \in M_{m\times n}(R)$ has a Smith normal form. Furthermore, if $D_1 = \text{Diag}(d_1, \ldots, d_r)$ and $D_2 = \text{Diag}(s_1, \ldots, s_r)$ are two Smith normal forms of A, then $d_i \sim s_i$ for all $i = 1, \ldots, r$.

Proof. Since R is a PIR, A has a Smith normal form by Theorem 15.9. If $D_1 = \text{Diag}(d_1, \ldots, d_r)$ and $D_2 = \text{Diag}(s_1, \ldots, s_r)$ are two Smith normal forms of A, then $D_1 \approx A \approx D_2$. In particular, $D_1 \approx D_2$. We then have $Rd_i = Rs_i$ for $i = 1, \ldots, r$ by Lemma 15.16. We must show $Rd_i = Rs_i$ implies d_i and s_i are associates in R.

Fix $i \in \{1, \ldots, r\}$. By Theorem 14.6, $R = R_1 \oplus \cdots \oplus R_p$ with each R_j either a PID or a special PIR. We have seen that $Z(R_j) \subseteq J(R_j)$ for all $j = 1, \ldots, p$. Let $d_i = (x_1, \ldots, x_p)$ and $s_i = (y_1, \ldots, y_p)$. Here $x_j, y_j \in R_j$ for all $j = 1, \ldots, p$. Since $Rd_i = Rs_i$, $R_jx_j = R_jy_j$ for each j. We claim x_j and y_j are associates in R_j.

Fix $j \in \{1, \ldots, p\}$. If $x_j = 0 = y_j$, then certainly x_j and y_j are associates in R_j. Hence, we can assume $x_j \neq 0 \neq y_j$. Since $R_jx_j = R_jy_j$, $ax_j = y_j$ and $by_j = x_j$ for some $a, b \in R_j$. Then $aby_j = y_j$. Therefore, $(ab - 1)y_j = 0$. Since $y_j \neq$

0, $ab - 1 \in Z(R_j) \subseteq J(R_j)$. In particular, $ab = (ab - 1) + 1 \in U(R_j)$. Hence, $a \in U(R_j)$, and $x_j \sim y_j$.

Since x_j is an associate of y_j for each $j = 1, \ldots, p$, we conclude that $d_i = (x_1, \ldots, x_p) \sim (y_1, \ldots, y_p) = s_i$ for all $i = 1, \ldots, r$. ∎

We next introduce the invariant factors of a matrix. Let R be a PIR, and let $A \in M_{m\times n}(R)$. If $A = O$, then any matrix equivalent to A is zero and there is nothing interesting to say. Hence, we will assume $A \neq O$. Since R is a PIR, A has a Smith normal form $D = \text{Diag}(d_1, \ldots, d_r)$ with $d_1 \mid d_2 \mid \cdots \mid d_r$. Since $A \neq O$, $d_1 \neq 0$. We have noted that $Rd_1 \supseteq Rd_2 \supseteq \cdots \supseteq Rd_r$. Hence, there exists an integer α with $1 \leq \alpha \leq r = \min\{m,n\}$ such that $d_1, \ldots, d_\alpha$ are nonzero elements of R and $d_{\alpha+1}, \ldots, d_r$ are zero. The sequence $\mathfrak{D}(A) = \{d_1, \ldots, d_\alpha\}$ is called a sequence of invariant factors of A. We have seen in Theorem 15.24 that the sequence $\mathfrak{D}$ is unique up to associates in R. Thus, if $D_2 = \text{Diag}(s_1, \ldots, s_r)$ is a second Smith normal form of A and $\mathfrak{D}' = \{s_1, \ldots, s_\alpha\}$ the corresponding sequence of invariant factors of A, then $d_i \sim s_i$ for all $i = 1, \ldots, \alpha$. This slight ambiguity is often ignored in the literature, and $\mathfrak{D} = \{d_1, \ldots, d_\alpha\}$ is called the sequence of invariant factors of A. Let us present what we have said here as a formal definition.

Definition 15.25 Let R be a PIR, and let $A \in M_{m\times n}(R)^*$. A sequence $\mathfrak{D} = \{d_1, \ldots, d_\alpha\}$ of elements from R is called a sequence of invariant factors of A if there exists a Smith normal form $D = \text{Diag}(e_1, \ldots, e_r)$ of A such that the following properties are satisfied:

(a) $1 \leq \alpha \leq r = \min\{m,n\}$
(b) $d_1 = e_1, \ldots, d_\alpha = e_\alpha$.
(c) $e_1, \ldots, e_\alpha$ are nonzero elements in R.
(d) $e_{\alpha+1} = \cdots = e_r = 0$.

In other words the nonzero entities on the diagonal of a Smith normal form of A will be called a sequence of invariant factors of A. Theorem 15.24 implies any two nonzero matrices in $M_{m\times n}(R)$ are equivalent if and only if they have the same sequence of invariant factors up to associates. It is sometimes convenient to extend the definition of invariant factors to the zero matrix O. We will let the empty set $\varnothing$ denote the sequence of invariant factors of O.

Now suppose R is a PID. Let $A \in M_{m\times n}(R)^*$. Suppose $\mathfrak{D}(A) = \{d_1, \ldots, d_\alpha\}$ is a sequence of invariant factors of A. Therefore, $1 \leq \alpha \leq r = \min\{m,n\}$, $d_1, \ldots, d_\alpha \in R^*$, and $d_1 \mid d_2 \mid \cdots \mid d_\alpha$. $A \approx D = \text{Diag}(d_1, \ldots, d_\alpha, 0, \ldots, 0)$ with $r - \alpha$ zeros appearing on the diagonal of D. Notice then that $\alpha = \text{rk}(A)$. Since R is a PID, R is a unique factorization domain. Hence, there exist units $\kappa_1, \ldots, \kappa_\alpha \in U(R)$ and primes (i.e., irreducible elements) $p_1, \ldots p_s \in R$ such that the following equations are true.

15.26
$$\begin{aligned} d_1 &= \kappa_1 p_1^{a_{11}} p_2^{a_{12}} \cdots p_s^{a_{1s}} \\ d_2 &= \kappa_2 p_1^{a_{21}} p_2^{a_{22}} \cdots p_s^{a_{2s}} \\ &\vdots \qquad\qquad \vdots \\ d_\alpha &= \kappa_\alpha p_1^{a_{\alpha 1}} p_2^{a_{\alpha 2}} \cdots p_s^{a_{\alpha s}} \end{aligned}$$

The a_{ij} appearing in equation 15.26 are all nonnegative integers. We also assume the primes $p_1, \ldots, p_s$ are nonassociates. Thus, p_i is not an associate of p_j whenever $i \neq j$. Some of the integers a_{ij} appearing in 15.26 may indeed be zero to allow for the fact that p_j does not occur in the factorization of d_i. We have already noted that $d_1 \mid d_2 \mid \cdots \mid d_\alpha$ implies the units in the sequence $\mathfrak{D} = \{d_1, \ldots, d_\alpha\}$ come first. If some d_i is a unit, then $a_{i1} = \cdots = a_{is} = 0$ for that i. Also notice that $d_1 \mid d_2 \mid \cdots \mid d_\alpha$ implies $a_{1j} \leq a_{2j} \leq \cdots \leq a_{\alpha j}$ for each $j = 1, \ldots, s$.

The factors in equation 15.26 of the form $p_j^{a_{ij}}$ with $a_{ij} > 0$ (if any) are called the elementary divisors of A. To be more specific, suppose at least one invariant factor $d_1, \ldots, d_\alpha$ of A is a nonunit in R. Then at least one integer a_{ij} is positive in the factorizations in equation 15.26. We can then form the following sequence

15.27
$$\mathfrak{E}(A) = \{p_1^{a_{11}}, \ldots, p_s^{a_{1s}}, \ldots, p_1^{a_{\alpha 1}}, \ldots, p_s^{a_{\alpha s}} \mid a_{ij} > 0\}$$

The notation here means $p_j^{a_{ij}}$ is listed as a term in $\mathfrak{E}(A)$ if and only if $a_{ij} > 0$. Thus, the sequence $\mathfrak{E}(A)$ is just a listing of the prime power factors $p_j^{a_{ij}}$ corresponding to those entries of the matrix (a_{ij}) $[\in M_{\alpha \times s}(\mathbb{Z})]$ which are nonzero. The sequence $\mathfrak{E}(A)$ is called a sequence of elementary divisors of A. Notice that $\mathfrak{E}(A)$ may have many repeated terms.

Suppose every invariant factor $d_1, \ldots, d_\alpha$ of A is a unit in R. Then we will set $\mathfrak{E}(A) = \varnothing$ (the empty sequence). Thus, $\mathfrak{E}(A) = \varnothing$ if and only if every invariant factor of A is a unit in R. Since the sequence $\mathfrak{D}(A) = \{d_1, \ldots d_\alpha\}$ of invariant factors of A is unique up to associates in R, it easily follows that the sequence $\mathfrak{E}(A)$ is unique up to associates in R. If $A, B \in M_{m \times n}(R)^*$ such that $A \approx B$, then Theorem 15.24 implies A and B have the same sequence of invariant factors and the same sequence of elementary divisors up to associates in R. Consider the following examples.

Example 15.28 Let $R = \mathbb{Z}$.

(a) Suppose

$$A \approx \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 12 & 0 & 0 & 0 \\ 0 & 0 & 36 & 0 & 0 \\ 0 & 0 & 0 & 5400 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \in M_{5\times 5}(\mathbb{Z})$$

Then $\mathcal{D}(A) = \{1,12,36,5400\}$ is a sequence of invariant factors of A. Factoring the invariant factors of A as in equation 15.26, we have

$$\begin{aligned} 1 &= 2^0 \times 3^0 \times 5^0 \\ 12 &= 2^2 \times 3^1 \times 5^0 \\ 36 &= 2^2 \times 3^2 \times 5^0 \\ 5400 &= 2^3 \times 3^3 \times 5^2 \end{aligned}$$

Thus, $\mathcal{E}(A) = \{2^2,3^1,2^2,3^2,2^3,3^3,5^2\}$ is a sequence of elementary divisors of A.

(b) Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{in } M_{2\times 2}(\mathbb{Z})$$

Then $\mathcal{D}(A) = \{1,1\}$ and $\mathcal{D}(B) = \{1\}$. Since $\mathcal{D}(A)$ is not equal to $\mathcal{D}(B)$ (up to associates), we conclude A is not equivalent to B. We could also have observed that equivalent matrices have the same rank. Therefore, A and B are not equivalent. However, A and B do have the same sequence of elementary divisors since $\mathcal{E}(A) = \mathcal{E}(B) = \varnothing$. Thus, if two matrices have the same sequence of elementary divisors, we cannot conclude these matrices are equivalent.

(c) Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix} \quad \text{in } M_{3\times 3}(\mathbb{Z})$$

Then $\mathcal{D}(A) = \{1,2,3\}$ and $\mathcal{D}(B) = \{1,2,5\}$, $\mathcal{E}(A) = \{2,3\}$, and $\mathcal{E}(B) = \{2,5\}$. Since $\mathcal{D}(A)$ is not equal to $\mathcal{D}(B)$ up to associates in $\mathbb{Z}$, we conclude A and B are not equivalent. These two matrices have the same number of invariant factors (as well as the same number of units among their invariant factors) but different sequences of elementary divisors. ∎

Example 15.28b shows that if two matrices have the same sequence of elementary divisors, we cannot conclude that these matrices are equivalent. However, that example worked for rather trivial reasons. Neither matrix had any elementary divisors and the number of invariant factors was different for each matrix. We can prove the following results.

Lemma 15.29 Let R be a PID. Let $A,B \in M_{m\times n}(R)^*$. Suppose A and B have the same number of terms in their sequences of invariant factors. If $\mathcal{E}(A) = \mathcal{E}(B)$ up to associates in R, then $A \approx B$.

Proof. Suppose $\mathcal{D}(A) = \{d_1, \ldots, d_\alpha\}$ and $\mathcal{D}(B) = \{d_1', \ldots, d_\alpha'\}$. If $\mathcal{E}(A) = \varnothing$, then every d_i is a unit in R. Since $\mathcal{E}(B) = \mathcal{E}(A) = \varnothing$, every d_i' is also

a unit in R. In particular, $\mathcal{D}(A) = \mathcal{D}(B)$ up to associates in R. Consequently, $A \approx B$.

Hence, we can assume some d_i is a nonunit in R. The lemma follows from the observation that $\mathcal{E}(A)$ determines the nonunits in the sequence $\mathcal{D}(A)$. Suppose the equations in 15.26 represent the factorizations of the invariant factors $d_1, \ldots, d_\alpha$. We have noted that $a_{1j} \leq a_{2j} \leq \cdots \leq a_{\alpha j}$ for all $j = 1, \ldots, s$. This implies d_α is the least common multiple of the elements in $\mathcal{E}(A)$, $d_{\alpha-1}$ is the least common multiple of the elements in $\mathcal{E}(A)$ with the factors for d_α removed, etc. Thus, the nonunits of $\mathcal{D}(A)$ are completely determined by the elements in the sequence $\mathcal{E}(A)$. If $\mathcal{E}(A) = \mathcal{E}(B)$ up to associates in R, then the nonunits in $\mathcal{D}(A)$ and $\mathcal{D}(B)$ are the same up to associates. Since $\mathcal{D}(A)$ has the same number of terms as $\mathcal{D}(B)$, we can now conclude that every term in $\mathcal{D}(A)$ is an associate of the corresponding term in $\mathcal{D}(B)$. Thus, $\mathcal{D}(A) = \mathcal{D}(B)$ up to associates in R and, consequently, $A \approx B$. ∎

We will have more to say concerning invariant factors and elementary divisors in the next chapter. We finish this chapter with some applications of Theorem 15.9.

Suppose R is a PIR. Let M be a finitely generated R-module. Let $\Gamma = \{m_1, \ldots, m_n\}$ be an R-module basis of M. As in Chapter 13, if $M = (0)$, we will take $n = 1$ and $\Gamma = \{0\}$. In any case, Γ determines a short exact sequence for M.

$$\textbf{15.30} \quad (0) \mapsto K \overset{\iota}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$$

As usual, $\varepsilon = \{\varepsilon_1, \ldots, \varepsilon_n\}$ denotes the canonical basis of R^n, f is given by $f(\varepsilon_i) = m_i$ for all $i = 1, \ldots, n$, and $K = \operatorname{Ker}(f)$. Suppose $\{\delta_1, \ldots, \delta_m\}$ is an R-module basis of K. For each $i = 1, \ldots, m$, let $\delta_i = \sum_{j=1}^{n} c_{ij}\varepsilon_j$. Then $C = (c_{ij}) \in M_{m\times n}(R)$ is a relations matrix for M. The matrix C determines an R-module homomorphism $g : M_{1\times m}(R) \mapsto K$ given by $g(\xi) = (\xi C)^t$. Hence,

$$\textbf{15.31} \quad M_{1\times m}(R) \overset{g}{\mapsto} R^n \overset{f}{\mapsto} M \mapsto (0)$$

is a presentation of M.

By Theorem 15.9, R is an elementary divisor ring. Thus, C has a Smith normal form. Hence, there exists invertible matrices $P \in \mathrm{Gl}(m,R)$ and $Q \in \mathrm{Gl}(n,R)$ such that $PCQ = D = \operatorname{Diag}(d_1, \ldots, d_r)$, a Smith normal form of C. In particular, $d_1 \mid d_2 \mid \cdots \mid d_r$ and $r = \min\{m,n\}$. We can modify the right exact sequence in 15.31 by defining two new R-module homomorphisms $g' : M_{1\times m}(R) \mapsto R^n$ and $f' : R^n \mapsto M$ as follows: $g'(\xi) = (\xi D)^t$ and $f'(\lambda) = f((Q^{-1})^t\lambda)$. We then have the following complex of R-modules.

15.32 $M_{1\times m}(R) \overset{g'}{\mapsto} R^n \overset{f'}{\mapsto} M \mapsto (0)$

We claim the sequence in 15.32 is right exact. Since Q is invertible, the map $\lambda \mapsto (Q^{-1})^t \lambda$ is an isomorphism of R^n to R^n. In particular, f' is a surjective R-module homomorphism. For any row vector $\xi \in M_{1\times m}(R)$, we have

$$\begin{aligned} f'g'(\xi) &= f'[(\xi D)^t] = f'[Q^tC^tP^t\xi^t] = f[(Q^{-1})^tQ^tC^tP^t\xi^t] = f[(\xi PC)^t] \\ &= f(g(\xi P)) = 0 \end{aligned}$$

since 15.31 is right exact. Thus, $\mathrm{Im}(g') \subseteq \mathrm{Ker}(f')$. Finally, suppose $\gamma \in \mathrm{Ker}(f')$. Then $0 = f'(\gamma) = f((Q^{-1})^t\gamma)$. Since the sequence in 15.31 is right exact, there exists a row vector $\alpha \in M_{1\times m}(R)$ such that $(\alpha C)^t = g(\alpha) = (Q^{-1})^t\gamma$. Since P is an invertible matrix, there exists a $\beta \in M_{1\times m}(R)$ such that $\alpha = \beta P$. Then $(Q^t)^{-1}\gamma = (Q^{-1})^t\gamma = C^t\alpha^t = C^tP^t\beta^t$. Therefore, $\gamma = Q^tC^tP^t\beta^t = g'(\beta)$. We have now shown $\mathrm{Im}(g') = \mathrm{Ker}(f')$. In particular, the sequence in 15.32 is right exact.

Since $D = \mathrm{Diag}(d_1, \ldots, d_r)$,

$$\mathrm{Im}(g') = R(d_1\varepsilon_1) \oplus R(d_2\varepsilon_2) \oplus \cdots \oplus R(d_r\varepsilon_r)$$

Since 15.32 is a right exact sequence, we have

$$\begin{aligned} M &\cong R^n/\mathrm{Ker}(f') = R^n/\mathrm{Im}(g') \cong R^n/(R(d_1\varepsilon_1) \oplus \cdots \oplus R(d_r\varepsilon_r)) \\ &\cong R/Rd_1 \oplus R/Rd_2 \oplus \cdots \oplus R/Rd_r \oplus R^{n-r} \end{aligned}$$

Thus, M is a finite direct sum of cyclic R-modules with descending annihilators: $Rd_1 \supseteq Rd_2 \supseteq \cdots \supseteq Rd_r \supseteq (0) \supseteq \cdots \supseteq (0)$. We have now proved the following theorem.

Theorem 15.33 Let R be a PIR and M a finitely generated R-module. Then M is a finite direct sum of cyclic modules: $M \cong R/Rx_1 \oplus R/Rx_2 \oplus \cdots \oplus R/Rx_n$ with descending annihilators $Rx_1 \supseteq Rx_2 \supseteq \cdots \supseteq Rx_n$. Furthermore, if no summand R/Rx_i is zero here, then this decomposition is unique. ∎

The uniqueness statement in Theorem 15.33 means the following: If $M \cong R/Ry_1 \oplus \cdots \oplus R/Ry_p$ with $Ry_1 \supseteq \cdots \supseteq Ry_p$ and no R/Ry_i zero, then $p = n$, and $R/Rx_i = R/Ry_i$ for all $i = 1, \ldots, n$. This follows from Lemma 15.13.

Theorem 15.33 is most often used when R is a PID. In this case, several observations are worth making here. Suppose M is a finitely generated module over a principal ideal domain R. Let $C \in M_{m\times n}(R)$ be a relations matrix for M. Let $D = \mathrm{Diag}(d_1, \ldots, d_r)$ be a Smith normal form of C. We have seen in the proof of Theorem 15.33 that $M \cong R/Rd_1 \oplus \cdots \oplus R/Rd_r \oplus R^{n-r}$. Here $r = \min\{m,n\}$. If $C = \mathrm{O}$, then $d_1 = \cdots = d_r = 0$, and M is a free R-module of rank n. Suppose $C \neq \mathrm{O}$. Let $\mathfrak{D}(C) = \{d_1, \ldots, d_\alpha\}$ be the sequence of invariant factors of C derived from D. Then $1 \leq \alpha \leq r$, and $d_{\alpha+1}, \ldots, d_r = 0$.

Therefore, we can write M as follows: $M \cong R/Rd_1 \oplus \cdots \oplus R/Rd_\alpha \oplus R^{n-\alpha}$. In particular, the torsion submodule of M is $\mathcal{T}(M) = R/Rd_1 \oplus \cdots \oplus R/Rd_\alpha$. $\mathcal{T}(M) = (0)$ precisely when $Rd_1 = \cdots = Rd_\alpha = R$. Thus, $\mathcal{T}(M) = (0)$ if and only if $\mathcal{E}(C) = \varnothing$. Notice that $\mathcal{T}(M)$ is always a direct summand of M since $\mathcal{T}(M) \oplus R^{n-\alpha} = M$. We have now proved the following theorem.

Theorem 15.34 Let R be a PID. Let M be a finitely generated R-module. Then

(a) M is torsion free if and only if $\mathcal{E}(C) = \varnothing$ for some relations matrix C of M.

(b) M is the direct sum of its torsion submodule $\mathcal{T}(M)$ and a free submodule whose rank is unique.

(c) M is a free R-module if and only if M is torsion free. ∎

In 15.34a, if $C = O$, then we set $\mathcal{E}(C) = \varnothing$. We can now give an easy proof of Exercise 7 in Chapter 14. We list this result as a corollary to Theorem 15.34.

Corollary 15.35 Let R be a PID. Any submodule of a finitely generated free R-module is itself a free R-module.

Proof. Suppose $M \subseteq R^n$. Since R^n is torsion free, M is torsion free. Therefore, M is free by Theorem 15.34c. ∎

There are two classical applications of Theorem 15.33. The first application occurs in the theory of abelian groups. Suppose G is a finitely generated abelian group. If we write the group action on G additively, that is, $g_1g_2 = g_1 + g_2$, then G is a finitely generated $\mathbb{Z}$-module. The scalar multiplication between $\mathbb{Z}$ and G is defined as follows: Let $n \in \mathbb{Z}$ and $g \in G$. If $n > 0$, then $ng = g + \cdots + g$ (n times). If $n = 0$, then $ng = 0$. If $n < 0$, then $ng = -(|n|g)$. Theorem 15.33 implies G is a finite direct sum of cyclic groups,

$$G \cong \mathbb{Z}/\mathbb{Z}d_1 \oplus \cdots \oplus \mathbb{Z}/\mathbb{Z}d_r \oplus \mathbb{Z}^{n-r}$$

where $d_1 \mid d_2 \mid \cdots \mid d_r$. In particular, if G is finite, then $G \cong \mathbb{Z}/\mathbb{Z}d_1 \oplus \cdots \oplus \mathbb{Z}/\mathbb{Z}d_\alpha$ with $d_1 \mid d_2 \mid \cdots \mid d_\alpha$ and $d_1, \ldots, d_\alpha \in \mathbb{Z}^*$. Thus, a finite, abelian group G is a finite direct sum, $G = G_1 \oplus \cdots \oplus G_\alpha$ of cyclic subgroups G_i of G with $\text{order}(G_1) \mid \text{order}(G_2) \mid \cdots \mid \text{order}(G_\alpha)$.

The second application of Theorem 15.33 occurs in the theory of canonical forms of matrices. The construction of a Frobenius normal form of an $n \times n$ matrix (entries from a field) will depend on Theorem 15.33. We will discuss this application in detail in the next chapter.

EXERCISES

1. In the proof of Theorem 15.8, we claimed $d_1 \mid x$ and $d_1 \mid y$ implies

$$\begin{bmatrix} d_1 & x \\ y & z \end{bmatrix} \approx \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}$$

   Give a proof of this assertion.
2. Let R be a commutative ring. Let $\mathfrak{A}$ be an ideal in R. For any $x \in R$, show $R/\mathfrak{A} : Rx \cong x(R/\mathfrak{A})$ as R-modules.
3. If $Rx = Ry$ in an arbitrary commutative ring R, can we conclude $x \sim y$?
4. Give a careful proof of the assertion: $A \approx B$ if and only if A and B have the same sequence of invariant factors up to associates in R.
5. Let R be a PID. Let $A \in M_{m\times n}(R)$. Suppose $\mathrm{rk}(A) = r > 0$. For each $i = 1, \ldots, r$, let Δ_i denote the greatest common divisor of all $i \times i$ minors of A. Show $\mathfrak{D}(A) = \{\Delta_1, \Delta_2\Delta_1^{-1}, \ldots, \Delta_r\,\Delta_{r-1}^{-1}\}$.
6. Prove the uniqueness statement in Theorem 15.34b.
7. What is the situation in Theorem 15.34 if we assume R is a PIR instead of a PID?
8. Let $R = \mathbb{Z}/6\mathbb{Z} = \{0,1,2,3,4,5\}$. Find a Smith normal form of the following matrices:

   (a) $\begin{bmatrix} 2 & 4 \\ 2 & 4 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & 4 \\ 4 & 1 & 2 \\ 4 & 4 & 3 \end{bmatrix}$

9. Compute $\mathfrak{D}(A)$ and $\mathfrak{E}(A)$ for

$$A = \begin{bmatrix} 2 & 2 & 2 \\ 1 & 2 & 1 \\ 3 & 4 & 9 \end{bmatrix} \in M_{3\times 3}(\mathbb{Z})$$

10. Compute $\mathfrak{D}(A)$ and $\mathfrak{E}(A)$ for

$$A = \begin{bmatrix} 1 & 2 & X & 1 \\ X & 4X & X^2 & 2X \\ X^3 & 2X^3 & X^2 + X^4 & X^3 \\ X + 1 & (X + 1)(X^2 + 2) & X + X^2 & (X + 1)(X^2 + 1) \end{bmatrix} \in M_{4\times 4}(\mathbb{Q}[X])$$

11. Let $R = \mathbb{Z}[i]$ where $i^2 = -1$. Set $M = R^3/K$ where K is generated by

$$\delta_1 = \begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix}, \quad \delta_2 = \begin{bmatrix} 1+i \\ 1-i \\ i \end{bmatrix}, \quad \delta_3 = \begin{bmatrix} 7+5i \\ 9 \\ 12+9i \end{bmatrix},$$

$$\delta_4 = \begin{bmatrix} -11-9i \\ -14-i \\ -18-17i \end{bmatrix}$$

Write M as a direct sum of cyclic R-modules, and compute $\mathfrak{T}(M)$.

12. Find all solutions in $\mathbb{Z}$ of the following system of Diophantine equations:

$$\begin{aligned} 7x - 13y + 12z &= 6 \\ x + 5y + 3z &= 8 \end{aligned}$$

Hint: Use the Smith normal form of

$$\begin{bmatrix} 7 & -13 & 12 \\ 1 & 5 & 3 \end{bmatrix}$$

13. Show $R = \mathbb{Z}[X]$ is not a PID. Show the ideal $\mathfrak{A} = (3, X+1) \subseteq R$ is not a direct sum of cyclic R-modules.
14. Suppose (R,δ) is a Euclidean domain (see [5, p. 148]). Let $A \in M_{m\times n}(R)$. Show that a Smith normal form for A can be obtained from A using only elementary row and column operations on A.
15. Suppose (R,δ) is a Euclidean domain. Let $A \in M_{n\times n}(R)$, and assume $\det(A) \neq 0$. Show there exists a $P \in \mathrm{Gl}(n,R)$ such that

$$PA = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

and $\delta(a_{jj}) > \delta(a_{ij})$ for all $j = 1, \ldots, n$ and $i = 1, \ldots, j-1$.

16. Let R be a PID. Let $A,B \in M_{n\times n}(R)$. Assume $\det(AB) \neq 0$. Let $D_1 = \mathrm{Diag}(d_1, \ldots, d_r)$, $D_2 = \mathrm{Diag}(e_1, \ldots, e_r)$, and $D_3 = \mathrm{Diag}(f_1, \ldots, f_r)$ be Smith normal forms for A, B, and AB, respectively. Show $d_i \mid f_i$ and $e_i \mid f_i$ for all $i = 1, \ldots, r$.

17. Let R be a PID. Let $\mathfrak{A}$ be a left ideal in $T = M_{n\times n}(R)$. Show $\mathfrak{A} = TA$ for some matrix $A \in T$.
18. List all nonisomorphic abelian groups of order 36. Explain your answer carefully.
19. This exercise is a sharper version of Corollary 15.35: Let R be a PID. Let M be an R-submodule of R^n. Show R^n has a free R-module basis $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_n\}$ such that the following properties are satisfied:
    (a) There exist an integer $r \leq n$ and elements $a_1, \ldots, a_r \in R$ such that $\{a_1\lambda_1, \ldots, a_r\lambda_r\}$ is a free R-module basis of M.
    (b) $a_1 \mid a_2 \mid \cdots \mid a_r$.

16

The Frobenius Normal Form of a Matrix

In this chapter F will denote a field and X an indeterminate over F. Suppose $A \in M_{n\times n}(F)$. We will abbreviate the matrix $XI_n - A$ by writing $X - A$. Thus, $X - A \in M_{n\times n}(F[X])$. Since $F[X]$ is a PID, the matrix $X - A$ has a Smith normal form by Theorem 15.9. We will use the Smith normal form of $X - A$ to construct a Frobenius normal form of A. We begin with the following theorem, which is true for any commutative ring R.

Theorem 16.1 Let A_1, A_2, B_1, and B_2 be $n\times n$ matrices in $M_{n\times n}(R)$. Set $M_1 = A_1X + B_1$ and $M_2 = A_2X + B_2$. Suppose A_2 is invertible. Then $M_1 \approx M_2$ in $M_{n\times n}(R[X])$ if and only if $PM_1Q = M_2$ for some $P,Q \in \mathrm{Gl}(n,R)$.

Proof. Since $M_{n\times n}(R) \subseteq M_{n\times n}(R[X])$ and $\mathrm{Gl}(n,R) \subseteq \mathrm{Gl}(n,R[X])$, the implication from right to left is obvious. Suppose $A_1X + B_1 \approx A_2X + B_2$ in $M_{n\times n}(R[X])$. Then there exist $U,V \in \mathrm{Gl}(n,R[X])$ such that $U(A_1X + B_1)V = A_2X + B_2$. We have seen in Lemma 7.18 that $M_{n\times n}(R[X]) \cong (M_{n\times n}(R))[X]$. Since A_2 is a unit in $M_{n\times n}(R)$, we can divide M_2 into both U and V by Theorem 7.2. Thus, there exist matrices R_1, $R_2 \in M_{n\times n}(R[X])$ and P, $Q \in M_{n\times n}(R)$ such that

16.2 $U = M_2R_2 + P$ and $V = R_1M_2 + Q$.

Then

$$\begin{aligned} M_2 &= UM_1V = (M_2R_2 + P)M_1V = PM_1V + M_2R_2M_1V \\ &= PM_1V + M_2R_2U^{-1}M_2 = PM_1(R_1M_2 + Q) + M_2R_2U^{-1}M_2 \\ &= PM_1Q + PM_1R_1M_2 + M_2R_2U^{-1}M_2 \\ &= PM_1Q + (U - M_2R_2)M_1R_1M_2 + M_2R_2U^{-1}M_2 \\ &= PM_1Q + UM_1R_1M_2 + M_2(-R_2M_1R_1 + R_2U^{-1})M_2 \\ &= PM_1Q + M_2(V^{-1}R_1 - R_2M_1R_1 + R_2U^{-1})M_2 \end{aligned}$$

Set $H = V^{-1}R_1 - R_2M_1R_1 + R_2U^{-1}$. Then

16.3 $M_2 = PM_1Q + M_2HM_2$.

We view equation 16.3 as a polynomial identity in X with coefficients in $M_{n\times n}(R)$. Now

$$M_2HM_2 = (A_2HA_2)X^2 + (B_2HA_2 + A_2HB_2)X + B_2HB_2$$

Substituting this expression in equation 16.3 and counting degrees in X implies $H = O$. Therefore, $M_2 = PM_1Q$. In other words, $A_2X + B_2 = (PA_1Q)X + PB_1Q$. Comparing coefficients of X, we get $PA_1Q = A_2$ and $B_2 = PB_1Q$. Since $A_2 \in \mathrm{Gl}(n,R)$, $P,Q \in \mathrm{Gl}(n,R)$. This completes the proof of Theorem 16.1. ∎

Theorem 16.1 has an important application to similar matrices. Recall two matrices B_1 and B_2 in $M_{n\times n}(R)$ are said to be similar if $P^{-1}B_1P = B_2$ for some $P \in \mathrm{Gl}(n,R)$. If B_1 and B_2 are similar, we will write $B_1 \tilde{s} B_2$. The reader can easily check that $\tilde{s}$ is an equivalence relation on $M_{n\times n}(R)$.

Corollary 16.4 Let $B_1, B_2 \in M_{n\times n}(R)$. Then $X - B_1 \approx X - B_2$ in $M_{n\times n}(R[X])$ if and only if $B_1 \tilde{s} B_2$ in $M_{n\times n}(R)$.

Proof. Suppose $X - B_1 \approx X - B_2$. Set $A_1 = A_2 = I_n$, and follow the proof of Theorem 16.1 to equation 16.3. Again, we get $H = O$ and

16.5 $X + B_2 = P(X + B_1)Q = (PQ)X + PB_1Q$

for some $P,Q \in M_{n\times n}(R)$. As before, we view equation 16.5 as an identity in X with coefficients in $M_{n\times n}(R)$. Comparing coefficients of X, we have $PQ = I_n$ and $PB_1Q = B_2$. Thus, $B_1 \tilde{s} B_2$ in $M_{n\times n}(R)$.

If $B_1 \tilde{s} B_2$, then clearly $X - B_1 \approx X - B_2$. ∎

We will use Corollary 16.4 at the end of this chapter when computing Frobenius normal forms. We now return to the setting of the first paragraph of this chapter. Suppose F is a field, and let $A \in M_{n\times n}(F)$. Then $X - A \in M_{n\times n}(F[X])$, and Theorem 15.9 implies $X - A$ has a Smith normal form D in $M_{n\times n}(F[X])$. Thus,

16.6 $X - A \approx D = \text{Diag}(d_1(X), d_2(X), \ldots, d_n(X))$

In equation 16.6, $d_1(X), \ldots, d_n(X)$ are polynomials in $F[X]$ with the property that $d_1(X) \mid d_2(X) \mid \cdots \mid d_n(X)$.

A couple of important observations about equation 16.6 should be mentioned here. Since $X - A \approx D$, there exist invertible matrices $P, Q \in \text{Gl}(n, F[X])$ for which $(X - A) = PDQ$. Since P and Q are invertible, $\det(P)$, $\det(Q) \in U(F[X]) = F^*$. Therefore, $C_A(X) = \det(X - A) = \det(P)\det(D)\det(Q) \sim d_1(X)d_2(X)\cdots d_n(X)$. Since $C_A(X)$ is a monic polynomial of degree n, $d_i(X) \neq 0$ for $i = 1, \ldots, n$. Furthermore, $\sum_{i=1}^{n} \partial(d_i) = n$. Thus, $\mathfrak{D} = \{d_1(X), \ldots, d_n(X)\}$ is a sequence of invariant factors of $X - A$, and at least one $d_i(X)$ in this sequence is a polynomial of positive degree in $F[X]$. Of course, some of the d_i could very well be units. Consider the following simple example.

Example 16.7 Let $F = \mathbb{Q}$, and

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix} \in M_{2\times 2}(\mathbb{Q})$$

Then

$$X - A = \begin{bmatrix} X - 1 & 0 \\ 1 & X - 2 \end{bmatrix} \approx \begin{bmatrix} 1 & 0 \\ 0 & (X - 1)(X - 2) \end{bmatrix}$$

Therefore, $\mathfrak{D}(X - A) = \{1, (X - 1)(X - 2)\}$, and $\mathfrak{E}(X - A) = \{X - 1, X - 2\}$. ∎

Let $\mathfrak{D}(X - A) = \{d_1(X), \ldots, d_n(X)\}$ be a sequence of invariant factors of $X - A$. Since some d_i must be a nonunit in $F[X]$, there exists a positive integer β such that $1 \leq \beta \leq n$, $d_1(X), \ldots, d_{\beta-1}(X)$ are constants in F^* and $d_\beta(X), \ldots, d_n(X)$ are polynomials of positive degree in $F[X]$. Let $\mathfrak{E}(X - A) = \{q_1(X), \ldots, q_l(X)\}$ be a sequence of elementary divisors of $X - A$. Since at least one $d_i(X)$ has positive degree, $\mathfrak{E}(X - A) \neq \varnothing$. Thus, $l \geq 1$. If the nonunits $d_\beta(X), \ldots, d_n(X)$ are factored into primes as in equation 15.26, then

$$\begin{aligned} \mathfrak{E}(X - A) &= \{q_1(X), \ldots, q_l(X)\} \\ &= \{p_1^{a_{\beta 1}}, \ldots, p_s^{a_{\beta s}}, \ldots, p_1^{a_{n1}}, \ldots, p_s^{a_{ns}} \mid a_{ij} > 0\} \end{aligned}$$

We observed in Chapter 15 that a sequence of invariant factors of a matrix is unique only up to associates in the ring. In particular, if $\mathfrak{D}(X - A) = \{d_1(X), \ldots, d_n(X)\}$, and $c_1, \ldots, c_n \in F^*$, then $\{c_1d_1(X), \ldots, c_nd_n(X)\}$ is also a sequence of invariant factors of $X - A$. Hence, we can always assume that $d_\beta(X), \ldots, d_n(X)$ are monic polynomials. We can then factor each $d_i(X)$ into irreducible polynomials in $F[X]$ which are monic. In particular, we can assume the elementary divisors $q_1(X), \ldots, q_l(X)$ of $X - A$ are all monic polynomials in $F[X]$. Thus, in what follows, we will assume $d_\beta(X), \ldots, d_n(X)$ and $q_1(X), \ldots, q_l(X)$ are all monic polynomials in $F[X]$.

For each $i = 1, \ldots, l$, let $n(i) = \partial(q_i(X))$. Then $n(i) \geq 1$ for each i. Since

$$\begin{aligned} C_A(X) &\sim d_1(X)d_2(X) \cdots d_n(X) \sim d_\beta(X)d_{\beta+1}(X) \cdots d_n(X) \\ &= q_1(X)q_2(X) \cdots q_l(X) \end{aligned}$$

we have $n(1) + \cdots + n(l) = n$.

Fix $i \in \{1, \ldots, l\}$. Suppose

$$q_i(X) = X^{n(i)} + a_{n(i)-1}X^{n(i)-1} + \cdots + a_1X + a_0$$

Recall the companion matrix, $\mathrm{Com}(q_i)$, of $q_i(X)$ is the following $n(i) \times n(i)$ matrix.

16.8 $$\mathrm{Com}(q_i) = \begin{bmatrix} 0 & 0 \cdots 0 & -a_0 \\ 1 & 0 \cdots 0 & -a_1 \\ 0 & 1 \cdots 0 & -a_2 \\ \cdot & \cdot \quad\quad \cdot & \cdot \\ \cdot & \cdot \quad\quad \cdot & \cdot \\ \cdot & \cdot \quad\quad \cdot & \cdot \\ 0 & 0 \cdots 1 & -a_{n(i)-1} \end{bmatrix}$$

We mentioned in Chapter 7 that the characteristic polynomial of $\mathrm{Com}(q_i(X))$ is precisely $q_i(X)$. Therefore, $C_{\mathrm{Com}(q_i(X))}(X) = q_i(X)$. Set $B_i(X) = (X) - \mathrm{Com}(q_i)$. Then

16.9 $$B_i(X) = \begin{bmatrix} X & 0 \cdots & 0 & a_0 \\ -1 & X \cdots & 0 & a_1 \\ 0 & -1 \ldots & 0 & a_2 \\ \cdot & \cdot \quad\quad & \cdot & \cdot \\ \cdot & \cdot \quad\quad & \cdot & \cdot \\ \cdot & \cdot \quad\quad & \cdot & \cdot \\ 0 & 0 \cdots -1 & X + a_{n(i)-1} \end{bmatrix} \in M_{n(i)\times n(i)}(F[X])$$

and $\det(B_i(X)) = q_i(X)$.

We next observe that the matrix $B_i(X)$ has precisely one elementary divisor $q_i(X)$. To see this, apply the following sequence of elementary row and column operators to $B_i(X)$:

$$B_i(X) \tilde{r} \begin{bmatrix} 1 & -X & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & -X & \cdots & 0 & -a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -X - a_\alpha \\ X & 0 & 0 & \cdots & 0 & a_0 \end{bmatrix} \quad \text{(here } \alpha = n(i) - 1\text{)}$$

$$\tilde{c} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -X & \cdots & 0 & -a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -X - a_\alpha \\ X & * & * & \cdots & * & * \end{bmatrix} \tilde{r} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -X & \cdots & 0 & -a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -X - a_\alpha \\ 0 & * & * & \cdots & * & * \end{bmatrix}$$

$$\tilde{r}\,\tilde{c} \cdots \sim \mathrm{Diag}(1, \ldots, 1, g(X))$$

Here $\tilde{r}$ and $\tilde{c}$ denote row and column operations and $*$ is an entry whose precise value is not important. Since row and column operations are performed on $B_i(X)$ by multiplying $B_i(X)$ on the left and right, respectively, with invertible matrices, we see $B_i(X) \approx \mathrm{Diag}(1, \ldots, 1, g(X))$ in $M_{n(i)\times n(i)}(F[X])$. Again we use the fact that invertible matrices in $M_{n(i)\times n(i)}(F[X])$ have determinants which are nonzero constants in F. Therefore,

$$q_i(X) = \det(B_i(X)) \sim \det(\mathrm{Diag}(1, \ldots, 1, g(X))) \sim g(X)$$

In particular, $\mathfrak{D}(B_i(X)) = \{1, \ldots, 1, q_i(X)\}$ and $\mathfrak{C}(B_i(X)) = \{q_i(X)\}$. Thus, $B_i(X)$ has precisely one elementary divisor $q_i(X)$. One final remark here. Since $q_i(X) \sim g(X)$, $B_i(X) \approx \mathrm{Diag}(1, \ldots, 1, q_i(X))$. Notice that there are $n(i) - 1$ ones on the diagonal of $\mathrm{Diag}(1, \ldots, 1, q_i(X))$.

Set $B(X) = \mathrm{Diag}(B_1(X), \ldots, B_l(X))$. Thus, $B(X)$ is a block diagonal matrix having $B_i(X)$ as its $n(i)\times n(i)$*th* block. Since $n(1) + n(2) + \cdots + n(l) = n$, $B(X) \in M_{n\times n}(F[X])$. The determinant of $B(X)$ is

$$\det(B(X)) = \prod_{i=1}^{l} \det(B_i(X)) = \prod_{i=1}^{l} q_i(X) \sim C_A(X)$$

Let

$$D_i = \mathrm{Diag}(1, \ldots, 1, q_i(X)) \in M_{n(i)\times n(i)}(F[X])$$

Since $B_i(X) \approx D_i$, $B \approx \text{Diag}(D_1, \ldots, D_l)$. We can always permute the diagonal entries of a square matrix Z by a similarity $P^{-1}ZP$. In particular, we have

16.10 $B(X) \approx \text{Diag}(1, \ldots, 1, q_1(X), q_2(X), \ldots, q_l(X))$ in $M_{n\times n}(F[X])$

In the equivalence relation in 16.10, there are $\sum_{i=1}^{l} (n(i) - 1) = n - l$ ones appearing on the diagonal of $\bar{D} = \text{Diag}(1, \ldots, 1, q_1(X), \ldots, q_l(X))$. It might be wise to point out here that $\bar{D}$ is not necessarily a Smith normal form since q_i need not divide q_{i+1}.

We can now argue the central point of this discussion.

16.11 $X - A \approx B(X)$.

In order to prove this equivalence, we need Exercise 5 of Chapter 15.

Lemma 16.12 Let R be a PID, and suppose $A \in M_{m\times n}(R)$ has rank r. For each $i = 1, \ldots, r$, let Δ_i denote the greatest common divisor of all $i \times i$ minors of A. Then

$$A \approx \text{Diag}(\Delta_1, \Delta_2\Delta_1^{-1}, \Delta_3\Delta_2^{-1}, \ldots, \Delta_r\Delta_{r-1}^{-1}, 0, \ldots, 0)$$

Before proving this lemma, let us say a few words about the notation here. We have observed in 4.13 that the rank of A is just the classical rank ($\text{rank}_{Q(R)}(A)$) of A when A is viewed as an $m \times n$ matrix with entires from $Q(R)$, the quotient field of R. Thus, $r = \text{rk}(A) = \max\{i \mid \Delta_i \neq 0\}$. In particular, $0 \leq r \leq \min\{m,n\}$, and there are precisely $\min\{m,n\} - r$ zeros appearing on the diagonal of $\text{Diag}(\Delta_1, \ldots, \Delta_r\Delta_{r-1}^{-1}, 0, \ldots, 0)$. We also have $\Delta_1 \mid \Delta_2 \mid \cdots \mid \Delta_r$ from Laplace's expansion.

Proof of Lemma 16.12. If $r = 0$, then $A = O$, and there is nothing to prove. Hence, we may assume $r > 0$. Suppose $P \in \text{Gl}(m,R)$ and $Q \in \text{Gl}(n,R)$. Then for any $t \geq 1$, $I_t(PAQ) = I_t(A)$ by Corollary 4.8. In particular, the greatest common divisor of the $t \times t$ minors of PAQ divides the greatest common divisor of the $t \times t$ minors of A and vice versa. Thus, if $A \approx A'$ in $M_{m\times n}(R)$, then the greatest common divisor of the $t \times t$ minors of A is the same (up to associates) as the greatest common divisor of the $t \times t$ minors of A'.

Since R is a PID, A has a Smith normal form by Theorem 15.9. Suppose $A \approx D = \text{Diag}(d_1, \ldots, d_p)$ with $p = \min\{m,n\}$ and $d_1 \mid d_2 \mid \cdots \mid d_p$. Since $\text{rk}(A) = r$, $d_{r+1} = \cdots = d_p = 0$ and $d_r \neq 0$. Our comments in the first paragraph of this proof imply $\Delta_1 \sim d_1$, $\Delta_2 \sim d_1d_2$, $\ldots$, $\Delta_r \sim d_1d_2, \ldots, d_r$. Thus,

$$\begin{aligned} A &\approx \text{Diag}(d_1, \ldots, d_r, 0, \ldots, 0) \\ &\approx \text{Diag}(\Delta_1, \Delta_2\Delta_1^{-1}, \ldots, \Delta_r\Delta_{r-1}^{-1}, 0, \ldots, 0) \end{aligned}$$

∎

We can now give a proof of the equivalence relation in 16.11. By Lemma 16.12, it suffices to show $I_t(X - A) = I_t(B(X))$ for all $t = 1, \ldots, n$. Since $B(X) \approx \tilde{D} = \text{Diag}(1, \ldots, 1, q_1(X), \ldots, q_l(X))$, it suffices to show $I_t(X - A) = I_t(\tilde{D})$ for all $t = 1, \ldots, n$.

Let $M = F[X]/(d_\beta(X)) \oplus \cdots \oplus F[X]/(d_n(X))$. Recall $d_\beta, \ldots, d_n$ are the nonunits in the sequence $\mathfrak{D}(X - A) = \{d_1, \ldots, d_n\}$. We have the following short exact sequence of $F[X]$-modules.

$$\textbf{16.13} \quad (0) \mapsto K \overset{\iota}{\mapsto} (F[X])^n \overset{f}{\mapsto} M \mapsto (0)$$

The $F[X]$-module homomorphism f is given by

$$f(\varepsilon_i) = \begin{cases} 0 & \text{if } i = 1, \ldots, \beta - 1 \\ (0, \ldots, 0, \underset{i - \beta + 1}{1 + (d_i(X))}, 0, \ldots, 0) & \text{if } i = \beta, \ldots, n \end{cases}$$

As usual, $\{\varepsilon_1, \ldots, \varepsilon_n\}$ is the canonical basis of $(F[X])^n$, $K = \text{Ker}(f)$, and ι is the inclusion map. Clearly, f is surjective. Hence the complex in 16.13 is exact. Since $d_1, \ldots, d_{\beta-1}$ are units in F^*, the reader can easily check that $K = \bigoplus_{i=1}^n F[X](d_i\varepsilon_i)$. Therefore, $D = \text{Diag}(d_1, \ldots, d_n)$ is a relations matrix for M.

We have listed a sequence of elementary divisors of $X - A$ as $\mathfrak{E}(X - A) = \{q_1(X), \ldots, q_l(X)\}$. The q_i here are powers of monic, irreducible polynomials in $F[X]$, and there exist positive integers $\alpha(1), \ldots, \alpha(n - \beta + 1)$ such that

$$\begin{aligned} d_\beta(X) &= q_1(X)q_2(X) \cdots q_{\alpha(1)}(X) \\ d_{\beta+1}(X) &= q_{\alpha(1)+1}(X) \cdots q_{\alpha(1)+\alpha(2)}(X), \ldots, \\ d_n(X) &= q_{\alpha(1) + \cdots + \alpha(n-\beta)+1}(X) \cdots q_{\alpha(1) + \cdots + \alpha(n-\beta+1)}(X) \end{aligned}$$

Notice that each set of factors

$$\{q_1, \ldots, q_{\alpha(1)}\}, \{q_{\alpha(1)+1}(X), \ldots, q_{\alpha(1)+\alpha(2)}(X)\},$$

etc. consists of pairwise relatively prime polynomials. In particular, Theorem 14.1 implies

$$F[X]/(d_\beta(X)) \cong \bigoplus_{i=1}^{\alpha(1)} F[X]/(q_i(X))$$

$$F[X]/(d_{\beta+1}(X)) \cong \bigoplus_{i=\alpha(1)+1}^{\alpha(1)+\alpha(2)} F[X]/(q_i(X))$$

etc. In particular,

$$M = \bigoplus_{i=\beta}^{n} F[X]/(d_i(X)) \cong \bigoplus_{i=1}^{l} F[X]/(q_i(X))$$

Suppose $g : \bigoplus_{i=1}^{l} F[X]/(q_i(X)) \mapsto M$ is the isomorphism given in the last paragraph. Set $m_i = g(0, \ldots, 0, \underset{F[X]q_i}{1}, 0, \ldots, 0)$ for $i = 1, \ldots, l$. Then $M = F[X]m_1 \oplus F[X]m_2 \oplus \cdots \oplus F[X]m_l$, and $\text{Ann}_{F[X]}(m_i) = F[X]q_i$

for each $i = 1, \ldots, l$. In particular, we have a second short exact sequence for M.

16.14 $(0) \mapsto K' \overset{\iota}{\mapsto} (F[X])^n \overset{f'}{\mapsto} M \mapsto (0)$

In 16.14, f' is given by the following formula:

$$f'(\varepsilon_i) = \begin{cases} 0 & \text{if } i = 1, \ldots, n-l \\ m_{i-(n-l)} & \text{if } i = n-l+1, \ldots, n \end{cases}$$

As usual, $K' = \operatorname{Ker}(f')$. Clearly, K' is generated by $\{\varepsilon_1, \ldots, \varepsilon_{n-l}, q_1(X)\varepsilon_{n-l+1}, \ldots, q_l(X)\varepsilon_n\}$. In particular, $\tilde{D}$ is a relations matrix for M.

We have seen in AlG in Chapter 13 that the Fitting ideals of M can be computed from any relations matrix of M. Since $X - A \approx \operatorname{Diag}(d_1(X), \ldots, d_n(X)) = D$, $I_k(X - A) = I_k(D)$ for any $k \in \mathbb{Z}$. Hence,

$$I_{n-k}(X - A) = I_{n-k}(D) = \mathfrak{F}_k(M) = I_{n-k}(\tilde{D}) \quad \text{for all } k \in \mathbb{Z}$$

In particular, $I_t(X - A) = I_t(\tilde{D})$ for all $t = 1, \ldots, n$. This completes the proof of the equivalence in 16.11.

We can now put all this material together to prove the following theorem.

Theorem 16.15 Let F be a field, and let $A \in M_{n\times n}(F)$. Suppose $\mathfrak{D}(X - A) = \{d_1(X), \ldots, d_n(X)\}$ and $\mathfrak{C}(X - A) = \{q_1(X), \ldots, q_l(X)\}$. Then

$$A \tilde{s} \operatorname{Diag}(\operatorname{Com}(q_1(X)), \operatorname{Com}(q_2(X)), \ldots, \operatorname{Com}(q_l(X)))$$

Proof. We have seen in 16.11 that $X - A \approx B(X)$ in $M_{n\times n}(F[X])$.

$$\begin{aligned} B(X) &= \operatorname{Diag}(X - \operatorname{Com}(q_1(X)), \ldots, X - \operatorname{Com}(q_l(X))) \\ &= X - \operatorname{Diag}(\operatorname{Com}(q_1(X)), \ldots, \operatorname{Com}(q_l(X))) \end{aligned}$$

Therefore, $X - A \approx X - \operatorname{Diag}(\operatorname{Com}(q_1), \ldots, \operatorname{Com}(q_l))$. It now follows from Corollary 16.4 that $A \tilde{s} \operatorname{Diag}(\operatorname{Com}(q_1), \ldots, \operatorname{Com}(q_l))$ in $M_{n\times n}(F)$. ∎

The block diagonal matrix $\operatorname{Diag}(\operatorname{Com}(q_1(X), \ldots, \operatorname{Com}(q_l(X))$ is called the Frobenius normal form of A. It is also called the rational normal form of A, or sometimes the first natural form of A.

EXERCISES

1. In the proof of Theorem 16.15, we claimed (since we can assume with no loss of generality that $d_\beta(X), \ldots, d_n(X)$ are monic polynomials) we can assume $q_1(X), \ldots, q_l(X)$ are monic polynomials. Give a proof of this fact.
2. Let $f_1(X), \ldots, f_n(X)$ be irreducible polynomials in $F[X]$. Set $p_1(X) = f_1^{\alpha(1)}, \ldots, p_n(X) = f_n^{\alpha(n)}$. Here $\alpha(1), \ldots, \alpha(n) > 0$. Show there exists

a matrix $A \in M_{s\times s}(F)$ for some s such that $\mathcal{C}(X - A) = \{p_1(X), \ldots, p_n(X)\}$.

3. In the short exact sequence given in 16.14, show that $\{\varepsilon_1, \ldots, \varepsilon_{n-l}, q_1(X)\varepsilon_{n-l+1}, \ldots, q_l(X)\varepsilon_n\}$ is an $F[X]$-module basis of K'.
4. Find a Frobenius normal form of the following matrices:

(a) $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in M_{2\times 2}(\mathbb{Q})$

(b) $A = \begin{bmatrix} \sin(\theta) & -\cos(\theta) \\ \cos(\theta) & \sin(\theta) \end{bmatrix} \in M_{2\times 2}(\mathbb{R})$ ($\mathbb{R}$ is the field of real numbers)

(c) $A = \begin{bmatrix} -1 & 7 & 0 \\ 0 & 2 & 0 \\ 0 & 3 & -1 \end{bmatrix} \in M_{3\times 3}(\mathbb{Q})$

(d) $A = \begin{bmatrix} 1 & 1 & 2 & 0 \\ -1 & 4 & 5 & -4 \\ 0 & -1 & -1 & 1 \\ 0 & 4 & 2 & -4 \end{bmatrix} \in M_{4\times 4}(\mathbb{Q})$

(e) $A = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ -1 & 3 & -1 & 0 \end{bmatrix} \in M_{4\times 4}(\mathbb{R})$

5. Suppose F is an algebraically closed field (e.g., $F = \mathbb{C}$). Let $A \in M_{n\times n}(F)$. What does the Frobenius normal form of A look like in this case? How does the Frobenius normal form of A compare to the Jordan canonical form of A?
6. Let $A,B \in M_{n\times n}(F)$. Show that A is similar to B if and only if they have the same Frobenius normal form.
7. Let $A \in M_{n\times n}(F)$. Use the material in this section to show $A \tilde{s} A^t$.
8. Let $A \in M_{n\times n}(F)$. Suppose $\mathcal{D}(X - A) = \{d_1(X), \ldots, d_n(X)\}$. Show $d_n(X)$ is the minimal poynomial of A (up to associates).
9. Suppose $F \subseteq K$ are fields. Let $A,B \in M_{n\times n}(F)$. Use the material in this chapter to show $A \tilde{s} B$ is in $M_{n\times n}(F)$ if and only if $A \tilde{s} B$ is in $M_{n\times n}(K)$.
10. Use the material in this chapter to derive the real Jordan canonical form of a matrix. Thus, if $A \in M_{n\times n}(\mathbb{R})^*$ show A is similar to a block diagonal

matrix $\mathrm{Diag}(D_1, \ldots, D_r)$ in which each block D_i can have one of two forms:

$$\begin{bmatrix} \lambda & & & & \\ 1 & \lambda & & & \\ & 1 & \ddots & & \\ & & & \ddots & \\ & & & 1 & \lambda \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} D & & & & \\ I & D & & & \\ & I & \ddots & & \\ & & & \ddots & D \\ & & & & I \end{bmatrix}$$

Here

$$\lambda \in \mathbb{R}, \qquad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \quad \text{some } a,b, \in \mathbb{R}$$

11. A matrix $A \in M_{n\times n}(F)$ is said to be reducible if $A \tilde{s} \mathrm{Diag}(A_1, A_2)$ where $\mathrm{size}(A_1)$, $\mathrm{size}(A_2) < \mathrm{size}(A)$. If A is not reducible, A is called an irreducible matrix.
    (a) Show A is irreducible if and only if $X - A$ has precisely one elementary divisor.
    (b) Show any $A \in M_{n\times n}(F)$ is similar to a direct sum of irreducible matrices: $A \tilde{s} \mathrm{Diag}(A_1, \ldots, A_r)$ with each A_i irreducible.
12. Let $A \in M_{n\times n}(F)$. Show A is similar to a diagonal matrix if and only if each elementary divisor of $X - A$ is linear.
13. Suppose $A,B \in M_{n\times n}(F)$ have the same irreducible characteristic polynomial $f(X) \in F[X]$. Show $A \tilde{s} B$.

17

Eigenvalues and Diagonalizing a Matrix

In this chapter we will discuss eigenvalues, eigenvectors, and their connection with diagonalizing a given matrix. As usual, R will denote an arbitrary commutative ring. The definitions of eigenvalues, eigenvectors, eigenspaces, and the spectrum of a matrix are the same as in the classical case when R is a field.

Definition 17.1 Let $A \in M_{n\times n}(R)$.

(a) An element $d \in R$ is called an eigenvalue of A if $A\xi = d\xi$ for some nonzero $\xi \in R^n$.

(b) $\mathcal{S}(A) = \{d \in R \mid d \text{ is an eigenvalue of } A\}$ is called the spectrum of A.

(c) A nonzero vector $\xi \in R^n$ is called an eigenvector of A if $A\xi = d\xi$ for some $d \in R$.

(d) Let $d \in \mathcal{S}(A)$. $E(d) = \{\xi \in R^n \mid A\xi = d\xi\}$ is called the eigenspace associated to d.

Eigenvalues (eigenvectors) are sometimes called characteristic values (characteristic vectors) in other texts. In this book, we will only use the names eigenvalue and eigenvector. The reader is already aware of the fact that a given matrix A may have no eigenvalues even if R is a field. For example,

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in M_{2\times 2}(\mathbb{R})$$

has no eigenvalues in $\mathbb{R}$, and no eigenvectors in $\mathbb{R}^2$.

Suppose $A \in M_{n\times n}(R)$ has an eigenvalue d. Then there exists a nonzero vector $\xi \in R^n$ such that $A\xi = d\xi$. The vector ξ will be called an eigenvector of A associated with d. Obviously, $\xi \in E(d)$. Clearly, $E(d) = \mathrm{NS}(dI_n - A)$. Thus, $E(d)$ is a nonzero R-submodule of R^n. The nonzero vectors in $E(d)$ are precisely the eigenvectors of A associated to d.

Our first lemma in this section generalizes the classical description of the spectrum of A.

Lemma 17.2 Let $A \in M_{n\times n}(R)$. Then the following sets are all the same.

(a) $\mathcal{S}(A)$

(b) $\{d \in R \mid \mathrm{NS}(dI_n - A) \neq (\mathrm{O})\}$

(c) $\{d \in R \mid C_A(d) \in Z(R)\}$

Proof. The fact that $\mathcal{S}(A) = \{d \in R \mid \mathrm{NS}(dI_n - A) \neq (\mathrm{O})\}$ is clear from Definition 17.1. Suppose $d \in \mathcal{S}(A)$. Then there exists a nonzero vector $\xi = (x_1, \ldots, x_n)^t \in R^n$ such that $(dI_n - A)\xi = \mathrm{O}$. Then $C_A(d)\xi = C_A(d)I_n\xi = \mathrm{adj}(dI_n - A)(dI_n - A)\xi = \mathrm{O}$. Since $\xi \neq \mathrm{O}$, some $x_i \neq 0$. Therefore, $C_A(d)x_i = 0$ implies $C_A(d) \in Z(R)$. In particular, $\mathcal{S}(A) \subseteq \{d \in R \mid C_A(d) \in Z(R)\}$.

Conversely, suppose $C_A(d) \in Z(R)$ for some $d \in R$. Since $C_A(d) = \det(dI_n - A)$, we have $\mathrm{rk}(dI_n - A) < n$ by 4.11e. It now follows from McCoy's theorem (Theorem 5.3) that $(dI_n - A)X = \mathrm{O}$ has a nontrivial solution $\xi \in R^n$. Thus, $A\xi = d\xi$. In particular, $d \in \mathcal{S}(A)$. Hence, $\{d \in R \mid C_A(d) \in Z(R)\} \in \mathcal{S}(\mathrm{A})$.

We can view the characteristic polynomial $C_A(X)$ of A as a polynomial function from R to R. The value of $C_A(X)$ on an element $z \in R$ is given by $C_A(z)$. Then Lemma 17.2 implies $\mathcal{S}(A) = C_A^{-1}(Z(R))$. Thus, if d is an eigenvalue of A, then $C_A(d)$ is a zero divisor in R. There is one important case in which d is actually a root of $C_A(X)$.

17.3 Suppose $d \in \mathcal{S}(A)$, and $A\xi = d\xi$ for some nonzero $\xi \in R^n$. If $\{\xi\}$ is linearly independent over R, then $C_A(d) = 0$.

The set $\{\xi\}$ is linearly independent over R if whenever $r\xi = \mathrm{O}$, then $r = 0$. Suppose $\{\xi\}$ is linearly independent over R. Since $C_A(d)\xi = C_A(d)I_n\xi = \mathrm{adj}(dI_n - A)(dI_n - A)\xi = \mathrm{O}$, we have $C_A(d) = 0$. This proves the assertion in 17.3.

If R is an integral domain, e.g., a field, then any nonzero vector in R^n is automatically linearly independent over R. In particular, 17.3 implies $\mathcal{S}(A) = \{d \in R \mid C_A(d) = 0\}$. Thus, we have the classical description of the eigenvalues of a matrix when R is a domain: The eigenvalues of A are precisely the roots of $C_A(X) = 0$ which lie in R.

We mention in passing that the converse of the assertion in 17.3 is not true in general. If $d \in \mathcal{S}(A)$ and $C_A(d) = 0$, it does not follow that an eigenvector of A associated to d is linearly independent over R. Consider the following example.

Example 17.4 Let $R = \mathbb{Z}/4\mathbb{Z} = \{0,1,2,3\}$. Set

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \in M_{2\times 2}(R)$$

$C_A(X) = X^2 + 2X + 1$, and 1 is a root of $C_A(X)$. By Lemma 17.2c, $1 \in \mathcal{S}(A)$. It is easy to check that

$$E(1) = NS(I_n - A) = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\}$$

None of the eigenvectors in $E(1)$ are linearly independent over R since $2E(1) = (O)$. ∎

Let us introduce the following notation for the roots of $C_A(X)$.

Definition 17.5 Let $A \in M_{n\times n}(R)$. Then $\mathcal{R}(A) = \{d \in R \mid C_A(d) = 0\}$.

Thus, $\mathcal{R}(A)$ is just the set of roots of $C_A(X)$ in R. If R is an integral domain, then $\mathcal{R}(A)$ contains at most n distinct roots. Of course, $C_A(X)$ may have no roots in R. For example, if

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in M_{2\times 2}(\mathbb{R})$$

then $C_A(X) = X^2 + 1$ has no roots in $\mathbb{R}$. Therefore, $\mathcal{S}(A) = \varnothing$. For an arbitrary commutative ring, $C_A(X)$ may have more than n distinct roots. Consider the following simple example.

Example 17.6 Let $X_1, X_2, \ldots$ be a countable collection of indeterminates over the field $\mathbb{Q}$. Set $R = \mathbb{Q}[X_1, X_2, \ldots]/\mathfrak{A}$ where $\mathfrak{A} = (X_1^2, X_2^2, \ldots)$. Let x_i denote the image of X_i in R.

Let $A = O \in M_{2\times 2}(R)$. Then $C_A(X) = X^2$, and $\mathcal{R}(A) \supseteq \{x_1, x_2, \ldots\}$. Thus, $C_A(X)$ has infinitely many zeros in R. ∎

At any rate, Lemma 17.2 implies that $\mathcal{R}(A) \subseteq \mathcal{S}(A)$. As it turns out, $\mathcal{R}(A)$ is the only important set of eigenvalues to consider when trying to decide if A is similar to a diagonal matrix. This follows from our next theorem.

Theorem 17.7 Let $A \in M_{n\times n}(R)$. A is similar to a diagonal matrix if and only if $\bigcup_{d\in\mathcal{R}(A)} E(d)$ contains a free R-module basis of R^n.

Proof. Suppose $A \sim D = \text{Diag}(d_1, \ldots, d_n)$ in $M_{n\times n}(R)$. Then there exists an invertible matrix $P \in \text{Gl}(n, R)$ such that $P^{-1}AP = D$. Let $P = (\delta_1 \mid \delta_2 \mid \cdots \mid \delta_n)$ be a column partition of P. Since $AP = (A\delta_1 \mid \cdots \mid A\delta_n)$, $PD = (d_1\delta_1 \mid \cdots \mid d_n\delta_n)$, and $AP = PD$, we conclude that $A\delta_i = d_i\delta_i$ for all $i = 1, \ldots, n$. Since P is invertible, $\{\delta_1, \ldots, \delta_n\}$ is a free R-module basis of R^n by Corollary 5.16. In particular, each set $\{\delta_i\}$ is linearly independent over R. Thus, $d_1, \ldots, d_n \in \mathcal{R}(A)$ by 17.3, and $\{\delta_1, \ldots, \delta_n\} \subseteq \bigcup_{d\in\mathcal{R}(A)} E(d)$. Therefore, $\bigcup_{d\in\mathcal{R}(A)} E(d)$ contains a free R-module basis of R^n.

Conversely, suppose $\bigcup_{d\in\mathcal{R}(A)} E(d) \supseteq \{\delta_1, \ldots, \delta_n\}$, a free R-module basis of R^n. Then each δ_i is an eigenvector of A associated to some eigenvalue $d_i \in \mathcal{R}(A)$. Set $P = (\delta_1 \mid \cdots \mid \delta_n)$. P is an invertible matrix by Corollary 5.16. Then

$$\begin{aligned} AP &= (A\delta_1 \mid \cdots \mid A\delta_n) = (d_1\delta_1 \mid \cdots \mid d_n\delta_n) \\ &= (\delta_1 \mid \cdots \mid \delta_n)\,\text{Diag}(d_1, \ldots, d_n) = P\,\text{Diag}(d_1, \ldots, d_n) \end{aligned}$$

Therefore, $P^{-1}AP = \text{Diag}(d_1, \ldots, d_n)$, and A is similar to a diagonal matrix. ∎

We will say a matrix $A \in M_{n\times n}(R)$ can be diagonalized if A is similar to a diagonal matrix in $M_{n\times n}(R)$. Theorem 17.7 says that, to decide whether A can be diagonalized, we need only consider the eigenspaces of A corresponding to the roots of $C_A(X)$. If these eigenspaces contain sufficiently many linearly independent vectors to span R^n, then A can be diagonalized. This is precisely the same situation as in the classical case when R is a field. Consider the following example.

Example 17.8 Let $R = \mathbb{Z}/6\mathbb{Z} = \{0,1,2,3,4,5\}$. Let

$$A = \begin{bmatrix} 2 & 3 \\ 4 & 3 \end{bmatrix} \in M_{2\times 2}(R)$$

A simple computation shows $C_A(X) = X^2 + X$. Then

$$\begin{aligned} C_A(0) &= 0 \in Z(R) & C_A(3) &= 0 \in Z(R) \\ C_A(1) &= 2 \in Z(R) & C_A(4) &= 2 \in Z(R) \\ C_A(2) &= 0 \in Z(R) & C_A(5) &= 0 \in Z(R) \end{aligned}$$

Therefore, $\mathcal{S}(A) = C_A^{-1}(Z(R)) = R$, and $\mathcal{R}(A) = \{0,2,3,5\}$. Notice that $C_A(X) \in R[X]$ is a monic polynomial of degree 2 having four distinct roots in R. Every element of R is an eigenvalue of A, but we need only examine four eigenspaces $E(0)$, $E(2)$, $E(3)$, and $E(5)$ to decide whether A can be diagonalized.

A simple calculation shows

$$E(0) = NS\begin{bmatrix}4 & 3\\ 2 & 3\end{bmatrix} = R\begin{bmatrix}3\\ 2\end{bmatrix}, \qquad E(2) = NS\begin{bmatrix}0 & 3\\ 2 & 5\end{bmatrix} = R\begin{bmatrix}1\\ 2\end{bmatrix}$$

$$E(3) = NS\begin{bmatrix}1 & 3\\ 2 & 0\end{bmatrix} = R\begin{bmatrix}3\\ 1\end{bmatrix}, \qquad E(5) = NS\begin{bmatrix}3 & 3\\ 2 & 2\end{bmatrix} = R\begin{bmatrix}1\\ 5\end{bmatrix}$$

Thus

$$\bigcup_{d \in \mathcal{R}(A)} E(d) = R\begin{bmatrix}3\\ 2\end{bmatrix} \cup R\begin{bmatrix}1\\ 2\end{bmatrix} \cup R\begin{bmatrix}3\\ 1\end{bmatrix} \cup R\begin{bmatrix}1\\ 5\end{bmatrix}$$

Notice that

$$\Gamma_1 = \left\{\begin{bmatrix}1\\ 5\end{bmatrix}, \begin{bmatrix}3\\ 2\end{bmatrix}\right\}$$

is a free R-module basis of R^2. Thus, Theorem 17.7 implies A is similar to a diagonal matrix. Set

$$P = \begin{bmatrix}1 & 3\\ 5 & 2\end{bmatrix}$$

Then

$$AP = \left[A\begin{bmatrix}1\\ 5\end{bmatrix} \middle| A\begin{bmatrix}3\\ 2\end{bmatrix}\right] = \begin{bmatrix}5 & 0\\ 1 & 0\end{bmatrix} = P\begin{bmatrix}5 & 0\\ 0 & 0\end{bmatrix}$$

Therefore, $P^{-1}AP = \text{Diag}(5,0)$. R is a PIR, and $\text{Diag}(5,0)$ is a Smith normal form of A.

Now

$$\Gamma_2 = \left\{\begin{bmatrix}1\\ 2\end{bmatrix}, \begin{bmatrix}3\\ 1\end{bmatrix}\right\}$$

is also a free R-module basis of R^2. Set

$$Q = \begin{bmatrix}1 & 3\\ 2 & 1\end{bmatrix}$$

Then

$$AQ = \left[A\begin{bmatrix}1\\2\end{bmatrix} \mid A\begin{bmatrix}3\\1\end{bmatrix}\right] = \begin{bmatrix}2 & 3\\4 & 3\end{bmatrix} = Q\begin{bmatrix}2 & 0\\0 & 3\end{bmatrix}$$

Therefore, $Q^{-1}AQ = \text{Diag}(2,3)$.

Thus, A is similar to at least two different diagonal matrices in $M_{2\times 2}(R)$. Notice that $\text{Diag}(2,3)$ is not a Smith normal form of A since 2 and 3 do not divide each other in R. ∎

Example 17.8 illustrates an important difference between fields and arbitrary commutative rings. If $A \tilde{s} B$ in $M_{n\times n}(R)$, then $XI_n - A \tilde{s} XI_n - B$ in $M_{n\times n}(R[X])$. In particular, $C_A(X) = C_B(X)$. Similar matrices have the same characteristic polynomial. Now suppose $R = F$, a field. If $A \tilde{s} \text{Diag}(d_1, \ldots, d_n)$ in $M_{n\times n}(F)$, then $C_A(X) = \prod_{i=1}^{n}(X - d_i)$. If $A \tilde{s} \text{Diag}(e_1, \ldots, e_n)$ in $M_{n\times n}(F)$, the same reasoning shows $C_A(X) = \prod_{i=1}^{n}(X - e_i)$. Since $F[X]$ is a unique factorization domain, we conclude that the sequence $\{e_1, \ldots, e_n\}$ is some permutation of $\{d_1, \ldots, d_n\}$. Thus, over a field, any diagonal matrix similar to A is unique up to a permutation of its diagonal entries.

This is certainly not the case over an arbitrary commutative ring. In Example 17.8, we saw $A \tilde{s} \text{Diag}(5,0)$, and $A \tilde{s} \text{Diag}(2,3)$ in $M_{2\times 2}(R)$. The sequence $\{2,3\}$ is not a permutation of the sequence $\{5,0\}$.

In Example 17.8, there are four eigenspaces of A which are of interest. The spaces $E(0)$, $E(2)$, $E(3)$, and $E(5)$ are free R-submodules of R^2 having bases

$$\left\{\begin{bmatrix}3\\2\end{bmatrix}\right\}, \quad \left\{\begin{bmatrix}1\\2\end{bmatrix}\right\}, \quad \left\{\begin{bmatrix}3\\1\end{bmatrix}\right\}, \quad \left\{\begin{bmatrix}1\\5\end{bmatrix}\right\}$$

respectively. Notice that the two vectors

$$\begin{bmatrix}3\\2\end{bmatrix} \text{ and } \begin{bmatrix}1\\2\end{bmatrix}$$

are not linearly independent over R since

$$3\begin{bmatrix}3\\2\end{bmatrix} + 3\begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} \text{ in } R^2$$

Thus, unlike the classical case in which R is a field, eigenvectors associated with distinct eigenvalues need not be linearly independent. If R is an integral domain, then we still have eigenvectors associated with distinct eigenvalues are linearly independent. In fact, we have the slightly more general statement.

Lemma 17.9 Let $A \in M_{n\times n}(R)$. Let $\Delta \subseteq \mathfrak{R}(A)$. Suppose Δ has the following property: If $d,d' \in \Delta$ and $d \neq d'$, then $d - d' \notin Z(R)$. Then $\sum_{d\in\Delta} E(d) = \oplus_{d\in\Delta} E(d)$.

Proof. We proceed by contradiction. Suppose the subspaces $\{E(d) \mid d \in \Delta\}$ are not independent in R^n. Then there exist distinct eigenvalues $d_1, \ldots, d_r \in \Delta$ and corresponding eigenvectors $\xi(d_j) \in E(d_j)$ for $j = 1, \ldots, r$ such that

$$\textbf{17.10} \quad \xi(d_1) + \xi(d_2) + \cdots + \xi(d_r) = O.$$

Among all such dependence relations, we can select one in which the fewest number of $\xi(d_i)$ appear. Suppose this relation is that given in equation 17.10. Then $\xi(d_j) \neq O$ for each $j = 1, \ldots, r$, and $r \geq 2$. Multiply equation 17.10 by d_1, and also apply A to the equation. We get the following two equations.

$$\textbf{17.11} \quad \begin{aligned} d_1\xi(d_1) + d_1\xi(d_2) + \cdots + d_1\xi(d_r) &= O \\ d_1\xi(d_1) + d_2\xi(d_2) + \cdots + d_r\xi(d_r) &= O \end{aligned}$$

Subtracting the two equations, we have

$$\textbf{17.12} \quad (d_1 - d_2)\xi(d_2) + \cdots + (d_1 - d_r)\xi(d_r) = O.$$

Each $(d_1 - d_i)\xi(d_i) \in E(d_i)$ for $i = 2, \ldots, r$. The minimality of the relation in equation 17.10 implies

$$(d_1 - d_2)\xi(d_2) = \cdots = (d_1 - d_r)\xi(d_r) = O$$

Now $\xi(d_2) \neq O$, and $d_1 - d_2 \notin Z(R)$ by hypothesis. Therefore, $(d_1 - d_2)\xi(d_2) = O$ is impossible. We conclude that no dependence relations are possible among the submodules $\{E(d) \mid d \in \Delta\}$. Therefore, $\sum_{d\in\Delta} E(d) = \oplus_{d\in\Delta} E(d)$. ∎

We note that the hypotheses of Lemma 17.9 are always satisfied if R is an integral domain. The principal application of Lemma 17.9 is the following corollary.

Corollary 17.13 Let $A \in M_{n\times n}(R)$. Suppose Δ is a subset of $\mathfrak{R}(A)$ with the following property: If $d,d' \in \Delta$ and $d \neq d'$, then $d - d' \notin Z(R)$. For each $d \in \Delta$, let $S(d)$ denote a maximal set of linearly independent vectors in $E(d)$. Then $\cup_{d\in\Delta} S(d)$ is a set of linearly independent eigenvectors in R^n. ∎

The proof of Corollary 17.13 is an immediate consequence of the fact that $\sum_{d\in\Delta} E(d) = \oplus_{d\in\Delta} E(d)$.

For the rest of this chapter, we will assume that R is an integral domain. Then Lemma 17.9 implies

17.14 $\sum_{d\in\mathcal{R}(A)} E(d) = \oplus_{d\in\mathcal{R}(A)} E(d)$ for any $A \in M_{n\times n}(R)$.

Theorem 17.7 and equation 17.14 suggest that a procedure for deciding when a given $A \in M_{n\times n}(R)$ (R an integral domain) can be diagonalized is as follows:

17.15

(a) Find a free R-module basis $S(d)$ of each eigenspace $E(d)$, $d \in \mathcal{R}(A)$.
(b) Show $\cup_{d\in\mathcal{R}(A)} S(d)$ is a free R-module basis of R^n.

Of course, for an arbitrary matrix $A \in M_{n\times n}(R)$, there is no reason to expect that either step (a) or (b) is possible. Consider the following two examples from [9].

Example 17.16 Let $R = \mathbb{Z}[\alpha]$ where $\alpha = \sqrt{-5}$. Then R is an integral domain contained in $\mathbb{C}$. Set

$$A = \begin{bmatrix} \alpha & 2 \\ 0 & 1 \end{bmatrix} \in M_{2\times 2}(R)$$

Since A is upper triangular, $\mathcal{S}(A) = \{\alpha,1\}$. Consider the eigenspace

$$E(1) = \mathrm{NS}\begin{bmatrix} 1-\alpha & -2 \\ 0 & 0 \end{bmatrix}$$

It is easy to check that

$$\xi_1 = \begin{bmatrix} 1+\alpha \\ 3 \end{bmatrix} \quad \text{and} \quad \xi_2 = \begin{bmatrix} 2 \\ 1-\alpha \end{bmatrix}$$

are nonzero vectors in $E(1)$. We claim $E(1)$ is not a free R-module.

To see this, suppose $E(1)$ is a free R-module. By passing to the quotient field $\mathbb{Q}(\alpha)$ of R, we see $E(1)$ must have rank one. Suppose

$$\lambda = \left\{\begin{bmatrix} x \\ y \end{bmatrix}\right\}$$

is a free R-module basis of $E(1)$. The integral domain R is quite familiar. $U(R) = \{1,-1\}$ and 2 and 3 are irreducible elements in R (see [5, pp. 141–142]). Since λ generates $E(1)$, $\xi_1 = c\lambda$ and $\xi_2 = d\lambda$ for some (unique) $c, d \in R$.

$\xi_1 = c\lambda$ implies $y = \pm 1$ or ± 3. $\xi_2 = d\lambda$ implies $x = \pm 1$ or ± 2. In particular, $x, y \in \mathbb{Q}^*$. But then

$$\lambda \in NS\begin{bmatrix} 1 - \alpha & -2 \\ 0 & 0 \end{bmatrix}$$

implies $\alpha \in \mathbb{Q}$. This is impossible. We conclude that the eigenspace $E(1)$ is not a free R-module. ∎

Example 17.16 shows that the first step (a) in 17.15 is not always possible. An eigenspace $E(d)$ of a given matrix A need not be a free R-module. Thus, no free basis $S(d)$ of $E(d)$ exists.

Example 17.17 Let $R = \mathbb{Z}$. Set

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \in M_{2\times 2}(\mathbb{Z})$$

Then $\mathcal{S}(A) = \{\pm 1\}$.

$$E(1) = NS\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \mathbb{Z}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$E(-1) = NS\begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} = \mathbb{Z}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Thus, both eigenspaces $E(1)$ and $E(-1)$ are free $\mathbb{Z}$-modules of rank one. Since

$$\det\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = -2 \notin U(\mathbb{Z})$$

it is easy to see that $S(1) \cup S(-1)$ cannot generate the $\mathbb{Z}$-module $\mathbb{Z}^2$ no matter how $S(1)$ or $S(-1)$ is chosen. Thus, $S(1) \cup S(-1)$ is not a free $\mathbb{Z}$-module basis of $\mathbb{Z}^2$ for any choice of $S(1)$ or $S(-1)$. ∎

Example 17.17 shows that step (b) in 17.15 is not always possible. Even though $\cup_{d\in\mathcal{R}(A)} S(d)$ consists of linearly independent vectors (Corollary 17.13), these vectors may not span R^n.

It turns out that (a) and (b) in 17.15 are the only impediments to diagonalizing a matrix. To be more specific, we have the following theorem of R. Richter and W. P. Wardlaw.

Theorem 17.18 Let R be an integral domain, and let $A \in M_{n\times n}(R)$. A is similar to a diagonal matrix in $M_{n\times n}(R)$ if and only if the following two conditions are satisfied:

(a) $E(d)$ is a free R-submodule of R^n for each $d \in \mathcal{R}(A)$.

(b) If $S(d)$ is a free R-module basis of $E(d)$ for each $d \in \mathcal{R}(A)$, then $\bigcup_{d\in\mathcal{R}(A)} S(d)$ is a free R-module basis of R^n.

Proof. If A satisfies conditions (a) and (b), then A is similar to a diagonal matrix in $M_{n\times n}(R)$ by Theorem 17.7. Suppose A is similar to a diagonal matrix $E \in M_{n\times n}(R)$. We can always permute the diagonal entries of E by passing to $Q^{-1}EQ$ for some suitable permutation matrix Q. Hence, there is a $P \in \mathrm{Gl}(n, R)$ such that

$$\textbf{17.19}\quad P^{-1}AP = D = \mathrm{Diag}(\underbrace{d_1, \ldots, d_1}_{n(1)}, \underbrace{d_2, \ldots, d_2}_{n(2)}, \ldots, \underbrace{d_k, \ldots, d_k}_{n(k)})$$

In equation 17.19, $d_1, \ldots, d_k$ are the distinct eigenvalues of A, and $n(1), \ldots, n(k)$ are the multiplicites of $d_1, \ldots, d_k$, respectively, in $C_A(X)$. In particular, $1 \leq k \leq n$, $1 \leq n(j) \leq n$ for each $j = 1, \ldots, k$, and $n(1) + n(2) + \cdots + n(k) = n$.

Since $P^{-1}AP = D$, $C_A(X) = \prod_{j=1}^{k}(X - d_j)^{n(j)}$ in $R[X]$. If $d \in \mathcal{R}(A)$, then $0 = C_A(d) = \prod_{j=1}^{k}(d - d_j)^{n(j)}$. Since R is an integral domain, we conclude $d = d_j$ for some $j \in \{1, \ldots, k\}$. Therefore, $\mathcal{R}(A) = \{d_1, \ldots, d_k\}$.

Suppose $P = (\delta_1 \mid \delta_2 \mid \cdots \mid \delta_n)$ is a column partition of P. Set

$$P_1 = (\delta_1 \mid \cdots \mid \delta_{n(1)}),\ P_2 = (\delta_{n(1)+1} \mid \cdots \mid \delta_{n(1)+n(2)}),\ \ldots,$$
$$P_k = (\delta_{n(1)+\cdots+n(k-1)+1} \mid \cdots \mid \delta_{n(1)+\cdots n(k)})$$

Then

$$P_1 \in M_{n\times n(1)}(R),\ P_2 \in M_{n\times n(2)}(R),\ \ldots,\ P_k \in M_{n\times n(k)}(R)$$

and $P = (P_1 \mid P_2 \mid \cdots \mid P_k)$.

Since $P^{-1}AP = D$,

$$(AP_1 \mid AP_2 \mid \cdots \mid AP_k) = AP = PD = (d_1P_1 \mid \cdots \mid d_kP_k)$$

Therefore, $CS(P_j) \subseteq E(d_j)$ for each $j = 1, \ldots, k$. Since P is invertible, $\{\delta_1, \ldots, \delta_n\}$ is a free R-module basis of R^n by Corollary 5.16. In particular, $R^n = E(d_1) + E(d_2) + \cdots + E(d_k)$. Since R is an integral domain, Lemma 17.9 implies this sum is direct. Thus, we have the following equation.

$$\textbf{17.20}\quad R^n = E(d_1) \oplus E(d_2) \oplus \cdots \oplus E(d_k).$$

We can now prove (a). Fix $i \in \{1, \ldots, k\}$. For notational convenience, relabel the columns of P_i as follows: $P_i = (\lambda_1 \mid \lambda_2 \mid \cdots \mid \lambda_{n(i)})$. Let $\xi \in E(d_i)$. Since $\{\delta_1, \ldots, \delta_n\}$ is a free R-module basis of R^n, we have $\xi = x_1\delta_1 + \cdots + x_n\delta_n$ for some $x_1, \ldots, x_n \in R$. Since $CS(P_j) \subseteq E(d_j)$ for each $j = 1, \ldots, k$, we have

$$\textbf{17.21} \quad \xi - x_{n(1)+\cdots+n(i-1)+1}\lambda_1 + \cdots + x_{n(1)+\cdots+n(i)}\lambda_{n(i)} \in \textstyle\sum_{j\neq i} E(d_j)$$

The vector

$$\xi - x_{n(1)+\cdots+n(i-1)+1}\lambda_1 + \cdots + x_{n(1)+\cdots+n(i)}\lambda_{n(i)}$$

lies in $E(d_i)$, and $E(d_i) \cap (\sum_{j\neq i} E(d_j)) = (0)$ by equation 17.20. Thus

$$\xi = x_{n(1)+\cdots+n(i-1)+1}\lambda_1 + \cdots + x_{n(1)+\cdots+n(i)}\lambda_{n(i)}$$

In other words, $E(d_i) = R\lambda_1 + \cdots + R\lambda_{n(i)}$. Since $\{\lambda_1, \ldots, \lambda_{n(i)}\} \subseteq \{\delta_1, \ldots, \delta_n\}$, $\lambda_1, \ldots, \lambda_{n(i)}$ are linearly independent over R. Therefore,

$$S(d_i) = \{\delta_{n(1)+\cdots+n(i-1)+1}, \ldots, \delta_{n(1)+\cdots+n(i)}\}$$

is a free R-module basis of $E(d_i)$. This proves (a).

$S(d_1) \cup \cdots \cup S(d_k) = \{\delta_1, \ldots, \delta_n\}$, a free R-module basis of R^n. If $L_1, \ldots, L_k$ are any free R-module bases of $E(d_1), \ldots, E(d_k)$ respectively, then $L_1 \cup \cdots \cup L_k$ is a free R-module basis of R^n. This follows immediately from equation 17.20. This proves (b). ∎

Theorem 17.18 implies that the procedure laid out in 17.15 is a good one to follow when trying to decide if a given matrix A can be diagonalized. If some eigenspace $E(d)$ of A is not a free R-module, then Theorem 17.18 implies A is not similar to a diagonal matrix. If the eigenspaces $\{E(d) \mid d \in \mathcal{R}(A)\}$ are all free R-modules, then A may be similar to a diagonal matrix. We then proceed with step (b) in 17.15. For each $d \in \mathcal{R}(A)$, find a free R-module basis $S(d)$ of $E(d)$. Since R is an integral domain, $\cup_{d\in\mathcal{R}(A)} S(d)$ is a linearly independent set of vectors in R^n (Corollary 17.13). If $\cup_{d\in\mathcal{R}(A)} S(d)$ fails to span R^n, then again Theorem 17.18 implies A is not similar to a diagonal matrix. If $\cup_{d\in\mathcal{R}(A)} S(d)$ is a free R-module basis of R^n, then Theorem 17.18 implies A is similar to a diagonal matrix. In fact, if $\cup_{d\in\mathcal{R}(A)} S(d) = \{\delta_1, \ldots, \delta_n\}$ and $A\delta_i = d_i\delta_i$ for all $i = 1, \ldots, n$, then we have seen in the proof of Theorem 17.7 that $A \tilde{s}\, \mathrm{Diag}(d_1, \ldots, d_n)$.

Now suppose R is a PID. We have observed in Corollary 15.35 that all submodules of R^n are free. This remark certainly applies to the eigenspaces $\{E(d) \mid d \in \mathcal{R}(A)\}$ of an $n \times n$ matrix $A \in M_{n\times n}(R)$. Thus, condition (a) in Theorem 17.18 is always satisfied when R is a PID. Notice this remark implies the ring $\mathbb{Z}[\alpha]$

in Example 17.16 is not a PID. This is easy to check directly since $\mathbb{Z}[\alpha]$ is not a unique factorization domain. Example 17.17 shows that condition (b) in Theorem 17.18 is not always satisfied even if R is a PID. Thus, the appropriate corollary to Theorem 17.18 when R is a PID is the following statement.

Corollary 17.22 Suppose R is a PID. Let $A \in M_{n\times n}(R)$. For each $d \in \mathcal{R}(A)$, let $S(d)$ be a free R-module basis of $E(d)$. Then A is similar to a diagonal matrix in $M_{n\times n}(R)$ if and only if $\cup_{d\in\mathcal{R}(A)} S(d)$ is a basis of the R-module R^n. ∎

We finish this section with an algorithm from [9] which allows us to compute a basis $S(d)$ of each eigenspace $E(d)$. We continue to assume R is a PID. Let $A \in M_{n\times n}(R)$, and let $d \in \mathcal{R}(A)$. Consider the matrix $C = dI_n - A \in M_{n\times n}(R)$. By Lemma 17.2, there exists a nonzero vector $\xi \in \mathrm{NS}(C)$. If $C = \mathrm{O}$, then $A = dI_n$, and we can take $S(d) = \{\varepsilon_1, \ldots, \varepsilon_n\}$, the canonical basis of R^n. Thus, in what follows, we will assume $C \neq \mathrm{O}$.

Since R is a PID, there exists an invertible matrix $Q \in \mathrm{Gl}(n, R)$ such that CQ has the following properties:

17.23 $CQ = (\delta_1 \mid \delta_2 \mid \cdots \mid \delta_p \mid \mathrm{O} \mid \cdots \mid \mathrm{O})$ is a column partition in which
(a) CQ is a lower triangular matrix
(b) $\delta_1, \ldots, \delta_p$ are linearly independent in R^n
(c) $p < n$

The proof of the assertions in 17.23 is similar to Gaussian elimination. We can multiply C on the right with the usual elementary matrices which interchange columns of C, multiply a column of C by a unit from R, and add a multiple of one column of C to another column of C. In order to produce (greatest) common divisors of entries in a given row of C, we may also have to multiply C on the right with block diagonal matrices of the following form:

17.24 $\mathrm{Diag}(I_{n(1)}, \ldots, G, \ldots, I_{n(r)})$.

Here if $d = \text{g.c.d.}(a, b)$, then $d = ax + by$ for some $x, y \in R$. Set $s = b/d$, $t = -a/d$, and

$$G = \begin{bmatrix} x & s \\ y & t \end{bmatrix}$$

Then $(a, b)G = (d, 0)$ and

$$\begin{bmatrix} -t & s \\ y & -x \end{bmatrix}\begin{bmatrix} x & s \\ y & t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Thus, the matrix listed in equation 17.24 is invertible. We use these types of matrices to find greatest common divisors of adjacent entries in a given row. We then use elementary column operations to produce zeros. A judicious choice of right multipliers of C produces a matrix CQ having properties (a) and (b) in 17.23. We must have $p < n$, since $\mathrm{NS}(C) \neq (\mathrm{O})$.

Instead of proving the above assertions, we give an example which illustrates the procedure. The interested reader can consult [5, Section 3.7] for more details.

Example 17.25 Let $R = \mathbb{Z}$. Suppose

$$C = \begin{bmatrix} 6 & 10 & 60 \\ 1 & 3 & 12 \\ 7 & 13 & 72 \end{bmatrix} \in M_{3\times 3}(\mathbb{Z})$$

Then

$$C = \begin{bmatrix} 6 & 10 & 60 \\ 1 & 3 & 12 \\ 7 & 13 & 72 \end{bmatrix} \begin{bmatrix} -3 & 5 & 0 \\ 2 & -3 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 60 \\ 3 & -4 & 12 \\ 5 & -4 & 72 \end{bmatrix}$$

$$Q_1$$

$$\begin{bmatrix} 2 & 0 & 60 \\ 3 & -4 & 12 \\ 5 & -4 & 72 \end{bmatrix} \begin{bmatrix} 1 & 0 & -30 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 3 & -4 & -78 \\ 5 & -4 & -78 \end{bmatrix}$$

$$E_1$$

$$\begin{bmatrix} 2 & 0 & 0 \\ 3 & -4 & -78 \\ 5 & -4 & -78 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 3 & 4 & 78 \\ 5 & 4 & 78 \end{bmatrix}$$

$$E_2$$

$$\begin{bmatrix} 2 & 0 & 0 \\ 3 & 4 & 78 \\ 5 & 4 & 78 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -19 & 39 \\ 0 & 1 & -2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 3 & 2 & 0 \\ 5 & 2 & 0 \end{bmatrix}$$

$$Q_2$$

Thus, if $Q = Q_1E_1E_2Q_2$, then $Q \in \mathrm{Gl}(3,\mathbb{Z})$, and

$$CQ = \begin{bmatrix} 2 & 0 & 0 \\ 3 & 2 & 0 \\ 5 & 2 & 0 \end{bmatrix} = (\delta_1 \mid \delta_2 \mid 0)$$

Here $\delta_1 = (2, 3, 5)^t$ and $\delta_2 = (0, 2, 2)^t \in \mathbb{Z}^3$. Notice that δ_1 and δ_2 are linearly independent in $\mathbb{Z}^3$. In this example, $p = 2$. ∎

Now suppose Q has been chosen such that $CQ = (\delta_1 \mid \cdots \mid \delta_p \mid \mathrm{O} \mid \cdots \mid \mathrm{O})$ satisfies the conditions in 17.23. Here $C = dI_n - A$ with $d \in \mathcal{R}(A)$. Let $Q = (\lambda_1 \mid \lambda_2 \mid \cdots \mid \lambda_n)$ be a column partition of Q. Remember $p < n$. Then we have

17.26 $S(d) = \{\lambda_{p+1}, \ldots, \lambda_n\}$ is a free R-module basis of $E(d)$.

Since $Q \in \mathrm{Gl}(n, R)$, $\{\lambda_{p+1}, \ldots, \lambda_n\}$ is a linearly independent set of vectors by Corollary 5.16. Since

$$(\delta_1 \mid \cdots \mid \delta_p \mid \mathrm{O} \mid \cdots \mid \mathrm{O}) = CQ = (C\lambda_1 \mid C\lambda_2 \mid \cdots \mid C\lambda_n)$$

we have

$\mathrm{O} = C\lambda_i = (dI_n - A)\lambda_i$ for $i = p + 1, \ldots, n$. Therefore, $\{\lambda_{p+1}, \ldots, \lambda_n\} \subseteq E(d)$. Let $\xi \in E(d)$. Since $\{\lambda_1, \ldots, \lambda_n\}$ is a free R-module basis of R^n, we have $\xi = x_1\lambda_1 + \cdots + x_n\lambda_n$ for some $x_1, \ldots, x_n \in R$. Then

$$\mathrm{O} = C\xi = x_1C\lambda_1 + \cdots + x_nC\lambda_n = x_1\delta_1 + \cdots + x_p\delta_p$$

Since $\{\delta_1, \ldots, \delta_p\}$ is a linearly independent set of vectors in R^n, $x_1 = \cdots = x_p = 0$. Therefore, $\xi = x_{p+1}\lambda_{p+1} + \cdots + x_n\lambda_n$. This completes the proof of the assertion in 17.26.

In summary, we have the following algorithm for computing a free R-module basis $S(d)$ of $E(d)$ when R is a PID.

17.27 Let R be a PID, $A \in M_{n\times n}(R)$, and $d \in \mathcal{R}(A)$. To compute a free R-module basis $S(d)$ of $E(d)$ carry out the following steps:

(a) Set $C = dI_n - A$. If $C = \mathrm{O}$, take $S(d) = \varepsilon$. If $C \neq \mathrm{O}$ go to (b).
(b) Find an invertible matrix $Q \in \mathrm{Gl}(n, R)$ such that $CQ = (\delta_1 \mid \cdots \mid \delta_p \mid \mathrm{O} \mid \cdots \mid \mathrm{O})$ is lower triangular, with $\delta_1, \ldots, \delta_p$ linearly independent over R $(p < n)$.
(c) Partition Q into columns $Q = (\lambda_1 \mid \cdots \mid \lambda_n)$.
(d) Then $S(d) = \{\lambda_{p+1}, \ldots, \lambda_n\}$.

Let us return to Example 17.25 for an illustration of this procedure.

Example 17.28 Let $R = \mathbb{Z}$. Suppose

$$A = \begin{bmatrix} -5 & -10 & -60 \\ -1 & -2 & -12 \\ -7 & -13 & -71 \end{bmatrix} \in M_{3\times 3}(\mathbb{Z})$$

Then $C_A(X) = X^3 + 78X^2 - 79X$, and $\mathcal{R}(A) = \{0,1,-79\}$. To find a free Z-module basis of $E(1)$, for example, we follow the algorithm given in 17.27.

(a) $$C = 1I_3 - A = \begin{bmatrix} 6 & 10 & 60 \\ 1 & 3 & 12 \\ 7 & 13 & 2 \end{bmatrix}$$

(b) Let

$$Q = Q_1E_1E_2Q_2 = \begin{bmatrix} -3 & 5 & -15 \\ 2 & 3 & -3 \\ 0 & -1 & 2 \end{bmatrix}$$

(notation as in Example 17.25)
Then

$$CQ = \begin{bmatrix} 2 & 0 & 0 \\ 3 & 2 & 0 \\ 5 & 2 & 0 \end{bmatrix}$$

Consequently, $p = 2$.

(c) $Q = (\delta_1 \mid \delta_2 \mid \delta_3)$ where

$$\delta_1 = \begin{bmatrix} -3 \\ 2 \\ 0 \end{bmatrix}, \quad \delta_2 = \begin{bmatrix} 5 \\ 3 \\ -1 \end{bmatrix}, \quad \delta_3 = \begin{bmatrix} -15 \\ -3 \\ 2 \end{bmatrix}$$

(d) $$S(1) = \left\{ \delta_3 = \begin{bmatrix} -15 \\ -3 \\ 2 \end{bmatrix} \right\}$$

∎

EXERCISES

1. If you have not done so already, verify all computations made in Example 17.8.
2. Find all diagonal matrices similar to A in Example 17.8.

3. Using the same techniques as in Example 17.8, decide if

$$A = \begin{bmatrix} 1 & 0 \\ 4 & 2 \end{bmatrix} \in M_{2\times 2}(\mathbb{Z}/6\mathbb{Z})$$

can be diagonalized.

4. Prove Corollary 17.13.
5. Is the matrix A given in Example 17.28 similar to a diagonal matrix? If so, find such a diagonal matrix.
6. Let $R = \mathbb{Z}$. Use the algorithms in 17.15 and 17.27 to find diagonal matrices similar to the following matrices:

(a) $A = \begin{bmatrix} 5 & 6 \\ -1 & 0 \end{bmatrix}$ (b) $A = \begin{vmatrix} -79 & -120 & -765 \\ -24 & -35 & -231 \\ 12 & 18 & 116 \end{vmatrix}$

7. Prove the claim in 17.23. The matrix CQ constructed there is called the Hermite normal form of C.
8. What does the Hermite normal form of C look like if R is a field?
9. Derive a corresponding Hermite form for nonsquare matrices $C \in M_{m\times n}(R)$ when R is a PID.
10. Let R be a PID. We have seen in Exercise 17 of Chapter 15 that left (and right) ideals in $M_{n\times n}(R)$ are principal. Use these ideas to develop a definition of greatest common left (and right) divisor of two elements $A_1, A_2 \in M_{n\times n}(R)$.
11. Use the Hermite normal form of a matrix to compute the greatest common left divisor of

$$A_1 = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 2 & -1 \\ 2 & 2 \end{bmatrix}$$

in $M_{2\times 2}(\mathbb{Z})$.

12. Let $R = \mathbb{Z}[\alpha]$ where $\alpha = \sqrt{-5}$. Show

$$A = \begin{bmatrix} 31 + 11\alpha & 45 + 15\alpha & 285 + 105\alpha \\ 6 + 18\alpha & 10 + 28\alpha & 57 + 177\alpha \\ -4 - 4\alpha & -6 - 6\alpha & -37 - 39\alpha \end{bmatrix} \in M_{3\times 3}(R)$$

can be diagonalized.

Appendix A
Partially Ordered Sets and Zorn's Lemma

Let A denote a set. Any subset of the crossed product $A \times A = \{\langle x, y\rangle \mid x, y \in A\}$ is called a relation on A. Suppose $R \subseteq A \times A$ is a relation on A. If $x, y \in A$ and $\langle x, y\rangle \in R$, then we say x relates to y. We will abbreviate the expression $\langle x, y\rangle \in R$ by writing $x \sim y$. Thus, $x \sim y$ (x relates to y) if and only if $\langle x, y\rangle \in R$. Although R does not appear in the symbols $x \sim y$, the definition of R will always be clear from the context in which these symbols are used. Often the symbol $\sim$ itself is called a relation on A.

Suppose R is a relation on some set A. As above, set $x \sim y$ if $\langle x, y\rangle \in R$.

Definition A.1 The relation $\sim$ is called a partial order on A if $\sim$ satisfies the following conditions:

(a) $x \sim x$ for all $x \in A$.
(b) For all $x, y \in A$, if $x \sim y$ and $y \sim x$, then $x = y$.
(c) For all $x, y, z \in A$, if $x \sim y$ and $y \sim z$, then $x \sim z$.

Any relation on A which satisfies condition (a) of Definition A.1 is said to be reflexive. A relation satisfying (b) is called an antisymmetric relation on A. Any relation satisfying (c) is called a transitive relation on A. Thus, a partial order on A is a reflexive, antisymmetric relation on A which is transitive.

There is certainly nothing unique about a partial order on A. A given set may admit many different partial orders. We will use the notation $(A, \sim)$ to indicate

a nonempty set A together with some partial order $\cdot$ on A. The ordered pair $(A, \cdot)$ will then be called a partially ordered set. Thus, if $(A, \cdot)$ is a partially ordered set, then A is a nonempty set and $\cdot$ is a partial order on A. Consider some simple examples.

Example A.2

(a) Let $A = \mathbb{Z}$, the integers. Let $\leq$ and $\geq$ denote the usual "less than or equal to" and "greater than or equal to" relations on $\mathbb{Z}$. Then $\leq$ and $\geq$ are two different partial orders on $\mathbb{Z}$. $(\mathbb{Z},\leq)$ and $(\mathbb{Z},\geq)$ are two examples of familiar partially ordered sets.

(b) Let $A = \mathbb{N}$, the natural numbers. Define a relation $\cdot$ on $\mathbb{N}$ by setting $x \cdot y$ if and only if $x \mid y$. It is easy to check that $(\mathbb{N},\cdot)$ is a partially ordered set.

(c) Let Γ denote a nonempty set. Let A be a nonempty subset of $\mathcal{P}(\Gamma)$, the set of all subsets of Γ. Thus, the elements of A are subsets of Γ. (For instance, Γ could be a commutative ring R and A the set of all ideals of R.) Then ordinary inclusion $\subseteq$ [$X \subseteq Y$ if every element of X is an element of Y] defines a partial ordering on the set A. Clearly, $\subseteq$ satisfies conditions (a) through (c) in Definition A.1. Thus $(A,\subseteq)$ is a partially ordered set. The partially ordered set $(A,\subseteq)$ is often referred to as A partially ordered by inclusion. ∎

Suppose $(A,\cdot)$ is a partially ordered set. If B is a nonempty subset of A, then the partial order $\cdot$ can be "restricted to B." By this we mean the following: Suppose $R \subseteq A \times A$ is the set which defines $\cdot$. Consider the (nonempty) set $R \cap (B \times B)$. Define a relation $\approx$ on B by setting $x \approx y$ if and only if $\langle x,y\rangle \in R \cap (B \times B)$. Since $\cdot$ is a partial order on A, it is easy to see that $\approx$ is a partial order on B. The relation $\approx$ is called the restriction of $\cdot$ to B. Henceforth, we will use the same symbol $\cdot$ to denote the restriction. Thus, if $(A,\cdot)$ is a partially ordered set and B is a nonempty subset of A, then $(B, \cdot)$ is a partially ordered set. In Example A.2c, the partially ordered set $(A, \subseteq)$ is obtained from $(\mathcal{P}(\Gamma), \subseteq)$ by restricting the partial order $\subseteq$ to A.

Let $(A,\cdot)$ be a partially ordered set. Two elements $x,y \in A$ are said to be comparable (relative to $\cdot$) if $x \cdot y$ or $y \cdot x$. A subset B of A is called a chain (or a linearly ordered subset of A) if any two elements in B are comparable. For instance, in Example A.2b, $B = \{2,4,8,16, \ldots\}$ is a chain in $(\mathbb{N}, \cdot)$.

Let B be a subset of a partially ordered set $(A,\cdot)$. An element $z \in A$ is called an upper bound of B if $x \cdot z$ for all $x \in B$. An element $z \in A$ is called a maximal element of $(A,\cdot)$ if there is no element $x \in \{y \in A \mid y \neq z\}$ such that $z \cdot x$. Another way to say this is as follows: $z \in A$ is a maximal element of $(A,\cdot)$ if whenever $z \cdot x$ (with $x \in A$), then $z = x$.

We can now state Zorn's lemma.

Lemma A.3 (Zorn's Lemma) Let $(A,\text{-})$ be a partially ordered set. If every chain in A has an upper bound (in A), then A contains a maximal element.

For a complete discussion of Zorn's lemma and how it relates to the other basic results in set theory, we refer the reader to reference [K] at the end of this appendix. In this book, our use of Zorn's lemma will be confined almost entirely to sets partially ordered by inclusion. For this situation, we have the following special case of Lemma A.3.

Corollary A.4 Suppose Γ is a nonempty set. Let A be a nonempty subset of $\mathcal{P}(\Gamma)$, the set of all subsets of Γ. Partially order A by inclusion $\subseteq$. If A contains the set theoretic union of the sets in any chain in $(A,\subseteq)$, then A contains a maximal element. ∎

As an application of Corollary A.4, let us argue that any ring T contains a maximal left ideal.

Example A.5 Let T be a ring. Thus, T is an associative ring with identity. We do not assume T is commutative. Let $\mathfrak{L}(T)$ denote the set of all proper left ideals of T. The set $\mathfrak{L}(T)$ is nonempty since $(0) \in \mathfrak{L}(T)$. We can partially order the ideals in $\mathfrak{L}(T)$ by inclusion. Thus, $(\mathfrak{L}(T),\subseteq)$ is a partially ordered set. In terms of the notation used in Corollary A.4, $\Gamma = \mathcal{P}(T)$ and $A = \mathfrak{L}(T)$.

Suppose B is a chain in $(\mathfrak{L}(T),\subseteq)$. Let $\mathfrak{A}$ denote the union of all the left ideals in B. Thus, $\mathfrak{A} = \cup\{\mathfrak{C} \mid \mathfrak{C} \in B\}$. We claim $\mathfrak{A}$ is a proper left ideal in T. To see this, let $x,y \in \mathfrak{A}$. Then there exist left ideals $\mathfrak{C}_1$ and $\mathfrak{C}_2$ in B such that $x \in \mathfrak{C}_1$ and $y \in \mathfrak{C}_2$. Since B is a chain, $\mathfrak{C}_1 \subseteq \mathfrak{C}_2$ or $\mathfrak{C}_2 \subseteq \mathfrak{C}_1$. Suppose $\mathfrak{C}_1 \subseteq \mathfrak{C}_2$. Then $x,y \in \mathfrak{C}_2$ and $x \pm y \in \mathfrak{C}_2 \subseteq \mathfrak{A}$. A similar proof shows $tx \in \mathfrak{A}$ for all $x \in \mathfrak{A}$ and $t \in T$. Thus, $\mathfrak{A}$ is a left ideal of T. If $\mathfrak{A}$ is not proper, then $\mathfrak{A} = T$. In particular, $1 \in \mathfrak{A}$. But then $1 \in \mathfrak{C}$ for some $\mathfrak{C} \in B$. Since $\mathfrak{C}$ is proper, this is impossible. We conclude that $\mathfrak{A} \in \mathfrak{L}(T)$.

We have now argued that the set theoretic union of the ideals in any chain of $(\mathfrak{L}(T),\subseteq)$ is another ideal in $\mathfrak{L}(T)$. By Corollary A.4, $(\mathfrak{L}(T),\subseteq)$ contains a maximal element $\mathfrak{M}$. Since, $\mathfrak{L}(T)$ is the set of all proper left ideals in T, $\mathfrak{M}$ is clearly a maximal left ideal of T. ∎

Reference

[K] John L. Kelly, *General Topology*, Van Nostrand, New York, 1955.

Appendix B
The Jacobson Radical

In this appendix, we will discuss the terminology appearing in Theorem 1.6 in the text. We will then give a proof of this theorem. Our treatment of the Jacobson radical is much the same as that found in [J]. The reader is urged to consult this reference for further reading on this subject.

As in Chapter 1 of this text, T will denote an associative ring with identity. We do not assume T is commutative. A left T-module M is an abelian group $(M,+)$ together with a function $f: T \times M \mapsto M$ (whose images we denote by $f(t,m) = tm$) which satisfies the following conditions:

B.1 (a) $t(m_1 + m_2) = tm_1 + tm_2$
(b) $(t_1 + t_2)m = t_1m + t_2m$
(c) $(t_1t_2)m = t_1(t_2m)$
(d) $1m = m$

These conditions are to hold for all $t,t_1,t_2 \in T$ and all $m,m_1, m_2 \in M$. We will refer to the elements of T as scalars and the elements of M as vectors. The function $f(t,m) = tm$ will be called scalar multiplication. A right T-module is an abelian group $(N,+)$ together with a function $g : N \times T \mapsto N$ (whose images we denote by $g(n,t) = nt$) which satisfies the same four conditions in B.1 but with the scalars $t \in T$ appearing on the right of the vectors $n \in N$. Thus, $(n_1 + n_2)t = n_1t + n_2t$ for all $n_1,n_2 \in N$ and all $t \in T$, etc.

For the time being, we will consider only left T-modules. However, let us take this opportunity to point out that if T is a commutative ring, then every left T-module is a right T-module and vice versa. For suppose M is a left T-module. Then M becomes a right T-module when scalar multiplication is defined as follows: $mt \doteq tm$. The only thing that needs checking here is whether this definition satisfies condition B.1c for right modules. Suppose $t_1, t_2 \in T$ and $m \in M$. Since T is commutative, we have

$$m(t_1t_2) = (t_1t_2)m = (t_2t_1)m = t_2(t_1m) = t_2(mt_1) = (mt_1)t_2$$

Thus, the left T-module M becomes a right T-module with scalar multiplication defined by $mt = tm$. Similar reasoning shows any right T-module is a left T-module via $tm \doteq mt$.

Suppose M is a left T-module. The annihilator, $\text{Ann}_T(M)$, of M is defined as follows: $\text{Ann}_T(M) = \{t \in T \mid tM = (0)\}$. Notice that $\text{Ann}_T(M)$ is always a two-sided ideal of T. In particular, the quotient ring $S = T/\text{Ann}_T(M)$ is well defined. It is easy to check that M is a left S-module when scalar multiplication is defined as follows: $(t + \text{Ann}_T(M))m = tm$ for all $t \in T$ and $m \in M$. Clearly, $\text{Ann}_S(M) = (0)$. A left T-module M is said to be faithful if $\text{Ann}_T(M) = (0)$. Thus, any left T-module M is a faithful $T/\text{Ann}_T(M)$-module. There is of course a similar discussion for right T-modules.

Now suppose $(M, +)$ is an arbitrary abelian group. Let $\mathbb{Z}$ denote the ring of integers. There is a natural function $g : \mathbb{Z} \times M \mapsto M$ which endows $(M, +)$ with the structure of a left (or right) $\mathbb{Z}$-module. The function g is defined as follows:

$$\textbf{B.2} \qquad g(x,m) = \begin{cases} m + m + \cdots + m & (x \text{ summands}) \quad \text{if } x \geq 1 \\ 0 & \text{if } x = 0 \\ -(m + m + \cdots + m) & (|x| \text{ summands}) \quad \text{if } x < 0 \end{cases}$$

Setting $xm = g(x,m)$ for all $x \in \mathbb{Z}$ and $m \in M$, the reader can easily verify that M is a left $\mathbb{Z}$-module with this scalar multiplication. We can now consider the set of all $\mathbb{Z}$-module homomorphisms from M to M, that is, $\text{Hom}_{\mathbb{Z}}(M,M)$. This set is an associative ring with identity. The sum and product of two $\mathbb{Z}$-module homomorphisms $h_1, h_2 \in \text{Hom}_{\mathbb{Z}}(M,M)$ is defined as follows:

$$\textbf{B.3} \qquad \begin{aligned} (h_1 + h_2)(m) &= h_1(m) + h_2(m) \quad \text{for all } m \in M \\ (h_1h_2)(m) &= h_1(h_2(m)) \quad \text{for all } m \in M \end{aligned}$$

The identity element of the ring $\text{Hom}_{\mathbb{Z}}(M,M)$ is the identity function $1_M : M \mapsto M$ given by $1_M(m) = m$ for all $m \in M$. Notice that $\text{Hom}_{\mathbb{Z}}(M,M)$ is usually a noncommutative ring since multiplication is performed by composing functions.

Definition B.4 Let T be a ring and $(M,+)$ an abelian group. Any ring homomorphism $\rho : T \mapsto \mathrm{Hom}_{\mathbb{Z}}(M,M)$ is called a representation of T.

Thus, to construct a representation of T, we must specify an abelian group $(M,+)$ and a ring homomorphism $\rho : T \mapsto \mathrm{Hom}_{\mathbb{Z}}(M,M)$. We remind the reader that $\rho : T \mapsto \mathrm{Hom}_{\mathbb{Z}}(M,M)$ is a ring homomorphism if $\rho(t_1 + t_2) = \rho(t_1) + \rho(t_2)$ and $\rho(t_1t_2) = \rho(t_1)\rho(t_2)$ for all $t_1,t_2 \in T$. We also have $\rho(1) = 1_M$; that is, ρ sends the identity of T to the identity of $\mathrm{Hom}_{\mathbb{Z}}(M,M)$.

If $\rho : T \mapsto \mathrm{Hom}_{\mathbb{Z}}(M,M)$ is a representation of T, then M becomes a left T-module when scalar multiplication is defined as follows: $tm = \rho(t)(m)$ [the value of $\rho(t)$ at m]. It is easy to check that the conditions in B.1 are all satisfied. The left T-module M constructed from ρ is called the left T-module corresponding to ρ. Thus, every representation $\rho : T \mapsto \mathrm{Hom}_{\mathbb{Z}}(M,M)$ determines a left T-module M corresponding to ρ.

The converse of this statement is also true. Suppose M is a left T-module. Then $(M,+)$ is an abelian group and hence a $\mathbb{Z}$-module. There is a natural representation $\rho : T \mapsto \mathrm{Hom}_{\mathbb{Z}}(M,M)$ given by $\rho(t) = \mu_t^L$. Here $\mu_t^L : M \mapsto M$ is the $\mathbb{Z}$-module homomorphism given by left multiplication by t; that is, $\mu_t^L(m) = tm$ for all $m \in M$. It is easy to check that $\rho(t) = \mu_t^L$ is a ring homomorphism from T to $\mathrm{Hom}_{\mathbb{Z}}(M,M)$. This representation will be called the regular representation of T corresponding to M.

We have seen that any representation of T determines a left T-module, and conversely any left T-module determines a representation of T. Notice if $\rho : T \mapsto \mathrm{Hom}_{\mathbb{Z}}(M,M)$ is a representation of T, and we view M as the corresponding left T-module, then $\mathrm{Ann}_T(M) = \mathrm{Ker}(\rho)$.

At this point, the reader might be wondering how right T-modules fit into this scheme. Right T-modules correspond to antirepresentations of T. If S and T are two rings, an antihomomorphism $f : S \mapsto T$ is an abelian group homomorphism from the underlying abelian group $(S,+)$ to the underlying group $(T,+)$ such that $f(1_S) = 1_T$ and $f(s_1s_2) = f(s_2)f(s_1)$ for all $s_1,s_2 \in S$. The transpose map $A \mapsto A^t$ is a good example of an antihomomorphism from $M_{n\times n}(F)$ to $M_{n\times n}(F)$. By an antirepresentation of T, we mean an antihomomorphism $\lambda : T \mapsto \mathrm{Hom}_{\mathbb{Z}}(M,M)$ for some abelian group $(M,+)$. Given such an antirepresentation λ, the reader can easily check that M becomes a right T-module when scalar multiplication is defined as follows: $mt = \lambda(t)(m)$ [the value of $\lambda(t)$ at m]. Conversely, if M is a right T-module, then $\lambda : T \mapsto \mathrm{Hom}_{\mathbb{Z}}(M,M)$ given by $\lambda(t) = \mu_t^R$ is an antirepresentation of T. Here $\mu_t^R : M \mapsto M$ denotes the $\mathbb{Z}$-module homomorphism given by right multiplication by t : $\mu_t^R(m) = mt$ for all $m \in M$.

We have seen that left (right) T-modules correspond to representations (antirepresentations) of T, and representations (antirepresentations) of T correspond to left (right) T-modules. For any statement about representations of T

and left T-modules there is a corresponding statement about antirepresentations of T and right T-modules. We will leave statements about antirepresentations of T and right T-modules to the reader until the end of this appendix.

Definition B.5 Let $\rho : T \mapsto \mathrm{Hom}_{\mathbb{Z}}(M,M)$ be a representation of T.
(a) ρ is irreducible if the corresponding T-module M is irreducible.
(b) ρ is faithful if $\mathrm{Ker}(\rho) = (0)$.

A left T-module M is irreducible if $M \neq (0)$ and (0) and M are the only left T-submodules of M. For example, if $\mathfrak{A}$ is a maximal left ideal in T, then the left T-module $T/\mathfrak{A}$ is clearly an irreducible left T-module. Notice that a representation ρ is faithful if and only if the annihilator of the corresponding left T-module M is zero. Thus, ρ is faithful if and only if the corresponding left T-module M is a faithful T-module.

Definition B.6 Let T be a ring and $\mathfrak{A}$ an ideal of T.
(a) T is primitive if T has a faithful, irreducible representation.
(b) $\mathfrak{A}$ is primitive if the ring $T/\mathfrak{A}$ is primitive.

As in Chapter 1, an ideal always means a two-sided ideal. Thus, if $\mathfrak{A}$ is an ideal of T, then $T/\mathfrak{A}$ is a well-defined ring. If T is a primitive ring, then T has a faithful, left T-module M which is irreducible. We already know some examples of primitive rings and primitive ideals. A field F is certainly a primitive ring since (up to isomorphism) F is the only irreducible F-module and $\mathrm{Ann}_F(F) = (0)$. It follows from this observation that any maximal ideal in a commutative ring R is a primitive ideal of R.

Definition B.7 Let $\mathfrak{A}$ be a left ideal of T. Then $\mathfrak{A} : T = \{t \in T \mid tT \subseteq \mathfrak{A}\}$.

If $\mathfrak{A}$ is a left ideal of T, then $\mathfrak{A}$ is a left T-submodule of the left T-module T. In particular, we can consider the quotient module $T/\mathfrak{A}$. Cleary, $\mathfrak{A} : T = \mathrm{Ann}_T(T/\mathfrak{A})$. In particular, $\mathfrak{A} : T$ is an ideal of T which is contained in $\mathfrak{A}$. It is easy to check that $\mathfrak{A} : T$ contains any ideal $\mathfrak{B}$ contained in $\mathfrak{A}$.

We will use the same notation employed in Chapter 1 and Appendix A. Thus, $\mathfrak{L}(T)$ will denote the set of all proper left ideals of T. A left ideal $\mathfrak{A} \in \mathfrak{L}(T)$ will be called a maximal left ideal if $\mathfrak{A}$ is a maximal element of the partially ordered set $(\mathfrak{L}(T),\subseteq)$. Thus, a left ideal $\mathfrak{A}$ of T is a maximal left ideal if $\mathfrak{A} \neq T$ and $\mathfrak{A}$ is not properly contained in any larger proper left ideal of T [i.e., $\mathfrak{A} \subseteq \mathfrak{B}$ with $\mathfrak{B} \in \mathfrak{L}(T) \Rightarrow \mathfrak{A} = \mathfrak{B}$]. We have already noted that if $\mathfrak{A}$ is a maximal left ideal of T, then the left T-module $T/\mathfrak{A}$ is irreducible. This follows directly from the definitions involved. Conversely, suppose M is an irreducible left T-module. Then $M \cong T/\mathfrak{A}$ for some maximal left ideal $\mathfrak{A}$ of T. To see this, let $m \in M -$

(0). Then $Tm = M$ since M is irreducible. The map $t \mapsto tm$ is a surjective homomorphism of left T-modules. Thus, the kernel $\mathfrak{A}$ of this map is a left ideal of T, and the isomorphism theorems imply $T/\mathfrak{A} \cong M$ as left T-modules. Since M is irreducible, $\mathfrak{A}$ is a maximal left ideal of T.

We can now give a more internal characterization of primitive rings.

Theorem B.8 A ring T is primitive if and only if T contains a maximal left ideal which contains no nonzero ideal of T.

Proof. Suppose T is a primitive ring. Then T has a faithful, irreducible left T-module M. We have seen in the discussion preceding this theorem that $M \cong T/\mathfrak{A}$ for some maximal left ideal $\mathfrak{A}$ of T. Since M is faithful, $(0) = \text{Ann}_T(M) = \text{Ann}_T(T/\mathfrak{A}) = \mathfrak{A} : T$. Since $\mathfrak{A} : T$ contains any ideal of T contained in $\mathfrak{A}$, we conclude $\mathfrak{A}$ contains no nonzero ideal of T.

Conversely, suppose $\mathfrak{A}$ is a maximal left ideal of T which contains no nonzero ideal of T. Since $\mathfrak{A}$ is maximal, $M = T/\mathfrak{A}$ is an irreducible left T-module. $\text{Ann}_T(M) = \mathfrak{A} : T \subseteq \mathfrak{A}$. Since $\mathfrak{A} : T$ is an ideal of T, we conclude $\mathfrak{A} : T = (0)$. Therefore, M is a faithful, irreducible, left T-module. In particular, T is a primitive ring. ∎

There are a couple of important corollaries to Theorem B.8.

Corollary B.9

(a) Any simple ring is primitive.

(b) A commutative ring R is primitive if and only R is a field.

Proof. (a) Recall that a ring T is simple if (0) and T are the only ideals of T. By Zorn's lemma (Example A.5 in Appendix A), T has at least one maximal left ideal $\mathfrak{A}$. Since T is simple, $\mathfrak{A}$ contains no nonzero ideal of T. Theorem B.8 implies T is primitive. (b) If R is a field, then R is a primitive ring. Suppose R is primitive. By Theorem B.8, R contains a maximal ideal $\mathfrak{A}$ which contains no nonzero ideal of R. Thus, $\mathfrak{A} = (0)$, and $R \cong R/(0)$ is a field. ∎

We have shown in Chapter 3 of this text that $M_{n\times n}(F)$ is a simple ring. Thus, $T = M_{n\times n}(F)$ is a non commutative, primitive ring when $n \geq 2$.

We can now introduce the Jacobson radical of T.

Definition B.10 The Jacobson radical $J(T)$ of T is the intersection of the kernels of all irreducible representations of T.

Since any irreducible representation $\rho : T \mapsto \text{Hom}_{\mathbb{Z}}(M,M)$ of T is given by an irreducible left T-module M, and $\text{Ker}(\rho) = \text{Ann}_T(M)$, the definition in B.10 can be rephrased as follows:

B.11 $J(T) = \cap \{\text{Ann}_T(M) \mid M \text{ an irreducible, left } T\text{-module}\}$.

The notation in equation B.11 means take the intersection of the annihilators of all irreducible, left T-modules. Since each $\text{Ann}_T(M)$ is a two-sided ideal of T, we observe that $J(T)$ is an ideal of T.

We are almost ready to start the proof of Theorem 1.6 We need one more lemma.

Lemma B.12 An ideal $\mathfrak{B}$ of T is primitive if and only if $\mathfrak{B} = \mathfrak{A} : T$ for some maximal left ideal $\mathfrak{A}$ of T.

Proof: Suppose $\mathfrak{B}$ is a primitive ideal of T. Then the ring $T/\mathfrak{B}$ is a primitive ring. Thus, $T/\mathfrak{B}$ has a faithful, irreducible left $T/\mathfrak{B}$-module M. The natural ring homomorphism $T \mapsto T/\mathfrak{B}$ endows M with the structure of a left T-module. Scalar multiplication is given by the formula $tm = (t + \mathfrak{B})m$ for all $t \in T$ and $m \in M$. It is easy to check that M is an irreducible left T-module with $\text{Ann}_T(M) = \mathfrak{B}$. Thus, $M \cong T/\mathfrak{A}$ for some maximal left ideal $\mathfrak{A}$ of T. In particular, $\mathfrak{B} = \text{Ann}_T(M) = \text{Ann}_T(T/\mathfrak{A}) = \mathfrak{A} : T$.

Suppose $\mathfrak{A}$ is a maximal left ideal of T. Set $\mathfrak{B} = \mathfrak{A} : T$. We have observed that $\mathfrak{B}$ is an ideal of T. Also, $\mathfrak{B} = \mathfrak{A} : T = \text{Ann}_T(T/\mathfrak{A})$. Since $\mathfrak{A}$ is a maximal left ideal of T, $T/\mathfrak{A}$ is an irreducible left T-module. Our remarks at the beginning of this appendix imply $T/\mathfrak{A}$ is a faithful, irreducible left $T/\text{Ann}_T(T/\mathfrak{A})$-module. Thus, $T/\mathfrak{A}$ is a faithful, irreducible left $T/\mathfrak{B}$-module. In other words, $\mathfrak{B}$ is a primitive ideal of T. ∎

We can now prove (c) and (a) of Theorem 1.6.

Theorem B.13 Let T be a ring.
(a) $J(T) = \cap \{\mathfrak{B} \mid \mathfrak{B} \text{ a primitive ideal of } T\}$.
(b) $J(T) = \cap \{\mathfrak{A} \mid \mathfrak{A} \text{ a maximal left ideal of } T\}$.

Proof. The proof of (a) follows directly from Lemma B.12. We have

$$\begin{aligned} J(T) &= \cap \{\text{Ann}_T(M) \mid M \text{ an irreducible, left } T\text{-module}\} \\ &= \cap \{\text{Ann}_T(T/\mathfrak{A}) \mid \mathfrak{A} \text{ a maximal left ideal of } T\} \\ &= \cap \{\mathfrak{A} : T \mid \mathfrak{A} \text{ a maximal left ideal of } T\} \\ &= \cap \{\mathfrak{B} \mid \mathfrak{B} \text{ a primitive ideal of } T\} \quad \text{(Lemma B.12)} \end{aligned}$$

(b) If M is any left T-module, then $\text{Ann}_T(M) = \cap_{m \in M^*} \text{Ann}_T(m)$. Here $\text{Ann}_T(m)$ is the left ideal of T defined by $\text{Ann}_T(m) = \{t \in T \mid tm = 0\}$. Now suppose M is an irreducible left T-module. Then for any $m \in M^*$, $M = Tm \cong T/\text{Ann}_T(m)$ as left T-modules. We conclude from this that $\text{Ann}_T(m)$ is a maximal left ideal of T for every $m \in M^*$. Therefore,

$$\begin{aligned} J(T) &= \cap \{\mathrm{Ann}_T(M) \mid M \text{ an irreducible, left } T\text{-module}\} \\ &= \cap \{\mathrm{Ann}_T(m) \mid m \in M^*, M \text{ an irreducible, left } T\text{-module}\} \\ &\supseteq \cap \{\mathfrak{A} \mid \mathfrak{A} \text{ a maximal left ideal of } T\} \end{aligned}$$

On the other hand, using (a), we have

$$\begin{aligned} &\cap \{\mathfrak{A} \mid \mathfrak{A} \text{ a maximal left ideal of } T\} \\ &\supseteq \cap \{\mathfrak{A} : T \mid \mathfrak{A} \text{ a maximal left ideal of } T\} \\ &= \cap \{\mathfrak{B} \mid \mathfrak{B} \text{ a primitive ideal of } T\} = J(T) \end{aligned}$$

Therefore,

$$J(T) = \cap \{\mathfrak{A} \mid \mathfrak{A} \text{ a maximal left ideal of } T\} \qquad \blacksquare$$

For the characterizations of $J(T)$ described in (d) and (f) of Theorem 1.6, we need the definition of the circle composition in T.

Definition B.14 For any $x,y \in T$, set $x \circ y = x + y - xy$.

The element $x \circ y$ is called the circle composition of x and y. The circle composition defines a function $\circ : T \times T \mapsto T$ given by $\circ(x,y) = x \circ y$. It is easy to check that $x \circ (y \circ z) = (x \circ y) \circ z$, and $0 \circ x = x \circ 0 = x$ for all $x,y,z \in T$. Thus, $(T,\circ,0)$ is an associative monoid with unit.

Definition B.15 Let T be a ring

(a) An element $x \in T$ is called left quasi-regular if $y \circ x = 0$ for some $y \in T$. An element $x \in T$ is called right quasi-regular if $x \circ z = 0$ for some $z \in T$. An element $x \in T$ is called quasi-regular if x is both left quasi-regular and right quasi-regular.

(b) A left ideal $\mathfrak{A}$ of T is called quasi-regular if every element in $\mathfrak{A}$ is left quasi-regular.

(c) A right ideal $\mathfrak{B}$ of T is called quasi-regular if every element in $\mathfrak{B}$ is right quasi-regular.

Several observations concerning these definitions are important here. We summarize these in our next lemma.

Lemma B.16 Let T be a ring.

(a) An element $x \in T$ is left (right) quasi-regular if and only if $1 - x$ has a left (right) inverse in T.

(b) An element $x \in T$ is quasi-regular if and only if $1 - x \in U(T)$.

(c) If $\mathfrak{A}$ is a left or right ideal of T which is quasi-regular, then every element in $\mathfrak{A}$ is quasi-regular.

Proof. (a) Let x be a left quasi-regular element of T. Then $0 = y \circ x = y + x - yx$ for some $y \in T$. This implies $(1 - y)(1 - x) = 1$. Thus, $1 - x$ has a left inverse in T. Conversely, suppose $z(1 - x) = 1$ for some $z \in T$. Write z as $z = 1 - y$ for some $y \in T$. Then $(1 - y)(1 - x) = 1$. This implies $y \circ x = 0$. Thus, x is left quasi-regular. A similar proof works for right quasi-regular elements of T.

(b) Suppose x is a quasi-regular element of T. Then $y \circ x = 0 = x \circ z$ for some $y,z \in T$. Since the circle composition is associative, we have $y = y \circ 0 = y \circ (x \circ z) = (y \circ x) \circ z = 0 \circ z = z$. Therefore, $y \circ x = 0 = x \circ y$. The same reasoning as in the proof of (a) then implies $(1 - y)(1 - x) = (1 - x)(1 - y) = 1$. Therefore, $1 - x \in U(T)$. Conversely, if $1 - x \in U(T)$, then $z(1 - x) = (1 - x)z = 1$ for some $z \in T$. Writing $z = 1 - y$, we have $y \circ x = 0 = x \circ y$. Therefore, x is quasi-regular.

(c) Suppose $\mathfrak{A}$ is a left ideal of T which is quasi-regular. Let $x \in \mathfrak{A}$. Then x is left quasi-regular by definition. Hence, $y + x - yx = y \circ x = 0$ for some $y \in T$. Since $\mathfrak{A}$ is a left ideal of T, this last equation implies $y \in \mathfrak{A}$. In particular, y is left quasi-regular. Suppose $z \circ y = 0$. As in the proof of (b), we then have $z = z \circ 0 = z \circ (y \circ x) = (z \circ y) \circ x = 0 \circ x = x$. Therefore, $x \circ y = y \circ x = 0$. Hence x is quasi-regular. A similar proof works for right quasi-regular ideals of T. ∎

We can now give a proof of (d) and (f) in Theorem 1.6.

Theorem B.17 Let T be a ring.

(a) $J(T)$ is a quasi-regular left ideal of T and contains every quasi-regular left ideal of T.

(b) $J(T) = \{z \in T \mid tz$ is left quasi-regular for all $t \in T\}$.

Proof. (a) Let $z \in J(T)$. Suppose z is not left quasi-regular. Then Lemma B.16a implies $1 - z$ has no left inverse in T. This is equivalent to saying $T(1 - z)$ is a proper left ideal of T. Using Zorn's lemma, we can find a maximal left ideal $\mathfrak{A}$ of T such that $T(1 - z) \subseteq \mathfrak{A}$. By Theorem B.13, $z \in \mathfrak{A}$. Therefore, $1 \in \mathfrak{A}$. This is clearly impossible. We conclude that z is left quasi-regular. Since z is an arbitrary element of $J(T)$, $J(T)$ is a left quasi-regular ideal of T.

Let $\mathfrak{B}$ be a left quasi-regular ideal of T. Suppose $\mathfrak{B}$ is not contained in $J(T)$. Theorem B.13 implies there exists a maximal left ideal $\mathfrak{A}$ of T such that $\mathfrak{B}$ is not contained in $\mathfrak{A}$. Since $\mathfrak{A}$ is maximal, and $\mathfrak{A} < \mathfrak{A} + \mathfrak{B}$, $\mathfrak{A} + \mathfrak{B} = T$. In particular, $1 = z + b$ for some $z \in \mathfrak{A}$ and $b \in \mathfrak{B}$. Since $\mathfrak{B}$ is a left quasi-regular ideal, b is left quasi-regular. Therefore, $z = 1 - b$ has a left inverse in T by Lemma B.16a. If $yz = 1$, then $1 \in \mathfrak{A}$ since $\mathfrak{A}$ is a left ideal of T. This is clearly impossible. We conclude that $\mathfrak{B} \subseteq J(T)$. This completes the proof of (a).

(b) The inclusion from left to right in (b) follows directly from (a). Suppose $z \in T$, and tz is left quasi-regular for all $t \in T$. Then Tz is a quasi-regular left ideal of T. By (a), $Tz \subseteq J(T)$. In particular, $z \in J(T)$. ∎

To derive the right characterizations of $J(T)$ given in (b), (e), and (g) of Theorem 1.6, we switch to right T-modules. Suppose we define a right Jacobson radical, $J^r(T)$, of T as follows:

$$J^r(T) = \cap \{\text{Ann}_T(M) \mid M \text{ an irreducible, right } T\text{-module}\}$$

Using right T-modules and antirepresentations of T, we can show $J^r(T)$ satisfies (b), (e), and (g) of Theorem 1.6. The arguments for these assertions are completely analogous to the proofs of (a), (d), and (f) for left ideals. Thus, $J^r(T)$ is a two-sided ideal of T, and every element in $J^r(T)$ is right quasi-regular. By Lemma B.16c, every element in $J^r(T)$ is left quasi-regular. Thus, $J^r(T) \subseteq J(T)$ by (d) of Theorem 1.6. If we reverse the roles of $J(T)$ and $J^r(T)$ in this reasoning, we get $J(T) \subseteq J^r(T)$. Thus, $J^r(T) = J(T)$, and, in particular, $J(T)$ satisfies (b), (e), and (g) of Theorem 1.6.

We have now proved all parts of Theorem 1.6. We finish this appendix with a few words about the commutative case. Suppose R is a commutative ring. Then there is no difference between a right ideal and a left ideal. Thus, $J(R)$ is the intersection of all maximal ideals of R. In a commutative ring R, an element $x \in R$ is a nonunit if and only if x is contained in some maximal ideal of R. This follows easily from Zorn's lemma. Thus, $z \in J(R)$ if and only if $1 + yz \in U(R)$ for all $y \in R$. This follows from Theorem 1.6f. Hence, we have the following descriptions of $J(R)$ when R is commutative.

Corollary B.18 Let R be a commutative ring.

(a) $J(R) = \cap \{\mathfrak{M} \mid \mathfrak{M} \text{ a maximal ideal of } R\}$.

(b) $J(R) = \{z \in R \mid 1 + yz \in U(R) \text{ for all } y \in R\}$. ∎

REFERENCE

[J] Nathan Jacobson, *Basic Algebra II*, second edition, W. H. Freeman, New York, 1989.

Appendix C
Elimination Theory and Bezout's Theorem

In this appendix, we discuss two of the principal applications of resultants, namely elimination theory and Bezout's theorem. Both of these topics come from judicious applications of Corollary 8.21 in the text.

Throughout this appendix, R will denote a commutative ring. There is one generalization of Corollary 8.21 which will be very convenient for what is to follow. Many authors allow a_0 or b_0 or both to be zero in Definition 8.12. Suppose $f(X)$ and $g(X)$ are polynomials in $R[X]$. Then f and g can be written as follows:

$$\textbf{C.1} \quad \begin{aligned} f(X) &= a_0X^n + a_1X^{n-1} + \cdots a_{n-1}X + a_n \\ g(X) &= b_0X^m + b_1X^{m-1} + \cdots + b_{m-1}X + b_m \end{aligned}$$

In the equations in C.1, $a_0, \ldots, a_n, b_0, \ldots, b_m \in R$. We make no further assumptions about a_0 and b_0, but we will always assume $m,n > 0$. If $a_0 = 0$, then the degree of $f(X)$ is less than n. When $f(X)$ is written as in equation C.1, the positive integer n is called the formal degree of $f(X)$. The constant a_0 is called the formal leading coefficient of $f(X)$. Thus, if $\partial(f)$ denotes the degree of $f(X)$, then $\partial(f)$ is less than or equal to the formal degree of f. Similar remarks can of course be made for $g(X)$.

Now suppose $f(X)$ and $g(X)$ are written as in equation C.1 with formal degrees n and m, respectively. In this case, we define the resultant $\mathcal{R}(f,g)$, of f and

g by the following formula: $\mathcal{R}(f,g) = S(a_0, \ldots, a_n, b_0, \ldots, b_m)$. Here $S(a_0, \ldots, a_n, b_0, \ldots, b_m)$ is the $[(n+m) \times (n+m)]$ determinant obtained from Sylvester's determinant $S(u_0, \ldots, u_n, v_0, \ldots, v_m)$ (Definition 8.11) by replacing $u_0, \ldots, u_n$ with $a_0, \ldots, a_n$ and $v_0, \ldots, v_m$ with $b_0, \ldots, b_m$. If a_0 and b_0 are not zero in C.1, then this new definition agrees with Definition 8.12. If $a_0 = b_0 = 0$, then obviously $\mathcal{R}(f,g) = 0$. Notice that our new definition of $\mathcal{R}(f,g)$ depends on the particular representation of f and g given in C.1.

In this appendix, we will adopt the new definition of $\mathcal{R}(f,g)$ given in the last paragraph. Slight modifications of some of the results in Chapter 8 are needed to accommodate this new definition. For example, Corollary 8.21 can now be stated as follows:

C.2 Suppose F is a field. Let $f(X), g(X) \in F[X]$ be written as in equation C.1 with $m,n > 0$. Then $\mathcal{R}(f,g) = 0$ if and only if either $a_0 = b_0 = 0$, or $f(X)$ and $g(X)$ have a common factor of positive degree in $F[X]$.

The proof of the assertion in C.2 readily follows from the proof of Corollary 8.21. See [J, Theorem 5.7].

Now suppose $R = \mathbb{Z}$ or $\mathbb{Z}/\mathbb{Z}\alpha$. Here $\mathbb{Z}$ denotes the ring of integers and α is a prime in $\mathbb{Z}$. Let $t_1, \ldots, t_p$ and X denote indeterminates over R. Consider the following r polynomials $f_1(X), \ldots, f_r(X)$ in $(R[t_1, \ldots, t_p])[X]$:

C.3
$$\begin{aligned} f_1(X) &= a_0^{(1)}X^{n(1)} + a_1^{(1)}X^{n(1)-1} + \cdots + a_{n(1)-1}^{(1)}X + a_{n(1)}^{(1)} \\ &\vdots \\ f_r(X) &= a_0^{(r)}X^{n(r)} + a_1^{(r)}X^{n(r)-1} + \cdots + a_{n(r)-1}^{(r)}X + a_{n(r)}^{(r)} \end{aligned}$$

In the equations in C.3, we assume the formal degree $n(i)$ of each $f_i(X)$ is a positive integer for each $i = 1, \ldots, r$. The coefficients $\{a_j^{(i)} \mid 0 \le j \le n(i); 1 \le i \le r\}$ appearing in C.3 are all polynomials in $R[t_1, \ldots, t_p]$.

Suppose F is any field containing R. Let $z_1, \ldots, z_p \in F$. Then there exists an R-algebra homomorphism $\sigma : R[t_1, \ldots, t_p] \mapsto F$ such that $\sigma(r) = r$ for all $r \in R$ and $\sigma(t_i) = z_i$ for $i = 1, \ldots, p$. Thus, if $g(t_1, \ldots, t_p) \in R[t_1, \ldots, t_p]$, then $\sigma(g) = g(z_1, \ldots, z_p) \in F$. The map σ induces an R-algebra homomorphism (which we continue to call σ) from $R[t_1, \ldots, t_p])[X]$ to $F[X]$ given by the following formula:

$$\sigma\left(\sum_{j=0}^{n} a_j(t_1, \ldots, t_p)X^j\right) = \sum_{j=0}^{n} a_j(z_1, \ldots, z_p)X^j$$

If

$$h(X) = \sum_{j=0}^{n} a_j(t_1, \ldots, t_p)X^j \in (R[t_1, \ldots, t_p])[X]$$

then $\sigma(h) = \sum_{j=0}^{n} a_j(z_1, \ldots, z_p)X^j \in F[X]$ will be denoted by $h^\sigma(X)$. The polynomial $h^\sigma(X)$ is called a specialization of $h(X)$ (obtained by replacing $t_1, \ldots, t_p$ with $z_1, \ldots, z_p$, respectively). The R-algebra map $\sigma : R[t_1, \ldots, t_p] \mapsto F$ is also called a specialization of $t_1, \ldots, t_p$. For the polynomials $f_1(X), \ldots, f_r(X)$ given in equation C.3, we have

$$\text{C.4} \quad \begin{aligned} f_1^\sigma(X) &= a_0^{(1)}(z_1, \ldots, z_p)X^{n(1)} + \cdots + a_{n(1)}^{(1)}(z_1, \ldots, z_p) \\ &\vdots \qquad\qquad\qquad\qquad \vdots \\ f_r^\sigma(X) &= a_0^{(r)}(z_1, \ldots, z_p)X^{n(r)} + \cdots + a_{n(r)}^{(r)}(z_1, \ldots, z_p) \end{aligned}$$

Notice that the formal degrees of the specializations $f_1^\sigma(X), \ldots, f_r^\sigma(X)$ are still $n(1), \ldots, n(r)$, respectively.

A natural question arises here. Given $f_1, \ldots, f_r$, is there some criterion based on the coefficients $\{a_j^{(i)} \mid 0 \le j \le n(i); 1 \le i \le r\}$ of the polynomials which will allow us to predict when the specialized polynomials $f_1^\sigma(X), \ldots, f_r^\sigma(X) \in F[X]$ have a common root in some extension field $F' \supseteq F$? This is equivalent to asking when the equations $f_1^\sigma(X) = \cdots f_r^\sigma(X) = 0$ have a common solution in some extension field $F' \supseteq F$. There is an answer to this question which uses resultants. The theorem is usually called Kronecker's method of elimination.

Theorem C.5 Let $R = \mathbb{Z}$ or $\mathbb{Z}/\mathbb{Z}\alpha$, α a prime in $\mathbb{Z}$. Suppose $t_1, \ldots, t_p$ and X are indeterminates over R. Let $f_1(X), \ldots, f_r(X)$ be the polynomials in $(R[t_1, \ldots, t_p])[X]$ listed in equation C.3. Then there exist polynomials $D_1, \ldots, D_h \in R[t_1, \ldots, t_p]$ with the following property: Let F be any field containing R and let $\sigma : R[t_1, \ldots, t_p] \mapsto F$ be any R-algebra homomorphism with $\sigma(t_i) = z_i$ for $i = 1, \ldots, p$. The equations $f_1^\sigma(X) = \cdots = f_r^\sigma(X) = 0$ have a common solution in some extension field $F' \supseteq F$, or the formal leading coefficients $a_0^{(1)}(z_1, \ldots, z_p), \ldots, a_0^{(r)}(z_1, \ldots, z_p)$ are all zero if and only if $D_1(z_1, \ldots, z_p) = \cdots = D_h(z_1, \ldots, z_p) = 0$.

Proof. Given $f_1, \ldots, f_r$ as in equation C.3, set $n = \max\{n(1), \ldots, n(r)\}$. Define $2r$ polynomials $l_1(X), \ldots, l_{2r}(X) \in (R[t_1, \ldots, t_p])[X]$ by the following equations:

$$\text{C.6} \quad \begin{aligned} l_1(X) &= X^{n-n(1)}f_1(X), \ldots, l_r(X) = X^{n-n(r)}f_r(X) \\ l_{r+1}(X) &= (X-1)^{n-n(1)}f_1(X), \ldots, l_{2r}(X) = (X-1)^{n-n(r)}f_r(X) \end{aligned}$$

Thus, each $f_i(X)$ is multiplied by $X^{n-n(i)}$ and $(X-1)^{n-n(i)}$ producing two new polynomials l_i and l_{r+i} for each $i = 1, \ldots, r$.

The polynomials $l_1(X), \ldots, l_{2r}(X)$ all have the same formal degree n. The formal leading coefficients of $l_1, \ldots, l_{2r}$ are $a_0^{(1)}, \ldots, a_0^{(r)}$, the formal leading coefficients of $f_1, \ldots, f_r$. If $\sigma : R[t_1, \ldots, t_p] \mapsto F$ is any specialization of $t_1, \ldots, t_p$, then the formal leading coefficients of $l_1^\sigma, \ldots, l_{2r}^\sigma$ are the same as the formal leading coefficients of $f_1^\sigma, \ldots, f_r^\sigma$. Since X and $X-1$ are relatively prime, it is easy to see that any common solution of the equations $l_1^\sigma(X) = \cdots = l_{2r}^\sigma(X) = 0$ (in some extension field $F' \supseteq F$) is a common solution of the equations $f_1^\sigma(X) = \cdots = f_r^\sigma(X) = 0$ and vice versa.

Suppose

$$l_i(X) = b_0^{(i)}X^n + b_1^{(i)}X^{n-1} + \cdots + b_{n-1}^{(i)}X + b_n^{(i)}$$

for each $i = 1, \ldots, 2r$. Then $\{b_0^{(1)}, \ldots, b_0^{(2r)}\} = \{a_0^{(1)}, \ldots, a_0^{(r)}\}$ in $R[t_1, \ldots, t_p]$.

Now let $u_1, \ldots, u_{2r}$ and $v_1, \ldots, v_{2r}$ be $4r$ indeterminates over $R[t_1, \ldots, t_p, X]$. Set $L_u = u_1l_1 + \cdots + u_{2r}l_{2r}$ and $L_v = v_1l_1 + \cdots + v_{2r}l_{2r}$. L_u and L_v are polynomials in $R_1[X]$ where $R_1 = R[t_1, \ldots, t_p, u_1, \ldots, u_{2r}, v_1, \ldots, v_{2r}]$. Since each polynomial $l_i(X)$ has the same formal degree n, the formal leading coefficient (with respect to X) of L_u is $\sum_{i=1}^{2r} b_0^{(i)}u_i$. Similarly, the formal leading coefficient of L_v is $\sum_{i=1}^{2r} b_0^{(i)}v_i$. The resultant, $\mathcal{R}_X(L_u, L_v)$, of L_u and L_v with respect to X is a polynomial in R_1.

We can now construct the polynomials $D_1, \ldots, D_h$ in the theorem. If $\mathcal{R}_X(L_u, L_v) = 0$, set $h = 1$ and $D_1 = 0$. Suppose $\mathcal{R}_X(L_u, L_v) \neq 0$. The polynomial ring R_1 is a free $R[t_1, \ldots, t_p]$-module with free basis the set of all monic monomials in the variables $u_1, \ldots, u_{2r}, v_1, \ldots, v_{2r}$. Since $\mathcal{R}_X(L_u, L_v) \neq 0$, there exist a finite number of nonzero polynomials $D_1, \ldots, D_h \in R[t_1, \ldots, t_p]^*$ and a finite number of (distinct) monic monomials

$$M_1(u_1, \ldots, u_{2r}, v_1, \ldots, v_{2r}), \ldots, M_h(u_1, \ldots, u_{2r}, v_1, \ldots, v_{2r})$$

such that

$$\textbf{C.7} \qquad \mathcal{R}_X(L_u, L_v) = \sum_{i=1}^{h} D_i(t_1, \ldots, t_p)M_i(u_i, \ldots, u_{2r}, v_1, \ldots, v_{2r})$$

The notation in equation C.7 includes the trivial case $\mathcal{R}_X(L_u, L_v) = 0$. In this case, take $h = 1$, $M_1 = 1$, and $D_1 = 0$.

Suppose $\sigma : R[t_1, \ldots, t_p] \mapsto F$ is a specialization of $t_1, \ldots, t_p$ with $\sigma(t_i) = z_i$ for $i = 1, \ldots, p$. In order to complete the proof of the theorem, we need the following claim:

Claim. The equations $l_1^{\sigma}(X) = \cdots = l_{2r}^{\sigma}(X) = 0$ have a common solution in some extention field $F' \supseteq F$, or $b_0^{(1)}(z_1, \ldots, z_p) = \cdots = b_0^{(2r)}(z_1, \ldots, z_p) = 0$ if and only if $D_1(z_1, \ldots, z_p) = \cdots = D_h(z_1, \ldots, z_p) = 0$.

Since the common solutions to $l_1^{\sigma}(X) = \cdots = l_{2r}^{\sigma}(X) = 0$ in $F' \supseteq F$ are the same as the common solutions to $f_1^{\sigma}(X) = \cdots = f_r^{\sigma}(X) = 0$, and the formal leading coefficients of $l_1^{\sigma}(X), \ldots, l_{2r}^{\sigma}(X) = 0$, are the same as the formal leading coefficients of $f_1^{\sigma}(X), \ldots, f_r^{\sigma}(X)$, the claim proves the theorem.

Before proving the claim, we observe that the equation given in C.7 is an identity in the variables $t_1, \ldots, t_p, u_1, \ldots, u_{2r}$ and $v_1, \ldots, v_{2r}$. In particular, the equation remains valid under any specialization of the form

$$t_1 \mapsto z_1, \ldots, t_p \mapsto z_p, u_1 \mapsto U_1, \ldots, u_{2r} \mapsto U_{2r},$$
$$v_1 \mapsto V_1, \ldots, v_{2r} \mapsto V_{2r}$$

Here $U_1, \ldots, U_{2r}, V_1, \ldots, V_{2r}$ are indeterminates over the ring $F[X]$.

To prove the claim, let us first suppose $D_1(z_1, \ldots, z_p) = \cdots = D_h(z_1, \ldots, z_p) = 0$. We can assume $b_0^{(i)}(z_1, \ldots, z_p) \neq 0$ for some $i \in \{1, \ldots, 2r\}$. Let $U_1, \ldots, U_{2r}$ and $V_1, \ldots, V_{2r}$ denote $4r$ indeterminates over $F[X]$. Set $L_U = U_1 l_1^{\sigma} + \cdots + U_{2r} l_{2r}^{\sigma}$ and $L_V = V_1 l_1^{\sigma} + \cdots + V_{2r} l_{2r}^{\sigma}$. Since $l_i^{\sigma}(X) \in F[X]$ for $i = 1, \ldots, 2r$, L_U and L_V are polynomials in $(F[U_1, \ldots, U_{2r}, V_1, \ldots, V_{2r}])[X]$. Since $b_0^{(i)}(z_1, \ldots, z_p) \neq 0$ for some i, the leading coefficients of L_U and L_V are not zero. Therefore, $\partial_X(L_U) = \partial_X(L_V) = n > 0$. Our comments in the preceding paragraph imply

$$\mathcal{R}_X(L_U, L_V) = \sum_{i=1}^{h} D_i(z_1, \ldots, z_p) M_i(U_1, \ldots, U_{2r}, V_1, \ldots, V_{2r}) = 0$$

Since $U_1, \ldots, U_{2r}, V_1, \ldots, V_{2r}$ are indeterminates over F, the ring $R_2 = F[U_1, \ldots, U_{2r}, V_1, \ldots, V_{2r}]$ is a unique factorization domain. Corollary 8.22 implies there exist a $g(X) \in R_2[X]$ with $\partial_X(g) > 0$ and polynomials $k_1(X), k_2(X) \in R_2[X]$ such that the following equations are valid:

C.8
$$U_1 l_1^{\sigma}(X) + \cdots + U_{2r} l_{2r}^{\sigma}(X) = g(X) k_1(X)$$
$$V_1 l_1^{\sigma}(X) + \cdots + V_{2r} l_{2r}^{\sigma}(X) = g(X) k_2(X)$$

Since $R_2[X]$ is an integral domain, $U_1, \ldots, U_{2r}, V_1, \ldots, V_{2r}$ are indeterminates over F, and $l_1^{\sigma}(X), \ldots, l_{2r}^{\sigma}(X) \in F[X]$, the equations in C.8 imply $g(X) \in F[X]$ and $g(X) \mid l_i^{\sigma}(X)$ in $F[X]$ for all i. In particular, $l_1^{\sigma}(X), \ldots, l_{2r}^{\sigma}(X)$ have a common factor $g(X)$ of positive degree in $F[X]$. It easily follows from this remark that the equations $l_1^{\sigma}(X) = \cdots = l_{2r}^{\sigma}(X) = 0$ have a common solution in some extension field $F' \supseteq F$ (e.g., F' any splitting field of $g(X)$ over F).

Conversely, suppose $b_0^{(1)}(z_1, \ldots, z_p) = \cdots = b_0^{(2r)}(z_1, \ldots, z_p) = 0$. Then the formal leading coefficients of L_U and L_V are zero. Therefore,

$$\begin{aligned}\mathcal{R}_X(L_U, L_V) &= \sum_{i=1}^{h} = D_i(z_1, \ldots, z_p) M_i(U_1, \ldots, U_{2r}, V_1 \ldots, V_{2r}) \\ &= 0\end{aligned}$$

Since the $M_i(U_1, \ldots, U_{2r}, V_1, \ldots, V_{2r})$, $i = 1, \ldots, h$, are distinct, monic monomials in the variables $U_1, \ldots, U_{2r}, V_1, \ldots, V_{2r}$, we conclude that $D_1(z_1, \ldots, z_p) = \cdots = D_h(z_1, \ldots, z_p) = 0$.

Suppose $b_0^{(i)}(z_1, \ldots, z_p) \neq 0$ for some $i \in \{1, \ldots, 2r\}$ and the equations $l_1^q(X) = \cdots = l_{2r}^q(X) = 0$ have a common solution α in some extension field $F' \supseteq F$. Then the formal leading coefficients of L_U and L_V are not zero. Thus, L_U and L_V are nonzero polynomials of degree n in X. Since $X - \alpha \mid l_i^q(X)$ in $F'[X]$ for all $i = 1, \ldots, 2r$, $X - \alpha \mid L_U$ and $X - \alpha \mid L_V$ in $(F'[U_1, \ldots, U_{2r}, V_1, \ldots, V_{2r}])[X]$. Corollary 8.22 implies

$$0 = \mathcal{R}_X(L_U, L_V) = \sum_{i=1}^{h} D_i(z_1, \ldots, z_p) M_i(U_1, \ldots, U_{2r}, V_1, \ldots, V_{2r})$$

As before, we conclude $D_1(z_1, \ldots, z_p) = \cdots = D_h(z_1, \ldots, z_p) = 0$. This completes the proof of the claim and, consequently, the proof of Theorem C.5. ∎

The polynomials $D_1, \ldots, D_h \in R[t_1, \ldots, t_p]$ constructed in equation C.7 are called a resultant system of $f_1, \ldots, f_r$. Let $R_1 = R[t_1, \ldots, t_p, u_1, \ldots, u_{2r}, v_1, \ldots, v_{2r}]$. If $A_1, \ldots, A_k$ are polynomials in $R_1[X]$, then we will let $(A_1, \ldots, A_k)$ denote the ideal in $R_1[X]$ generated by $A_1, \ldots, A_k$. Theorem 8.31 implies

$$\mathcal{R}_X(L_u, L_v) \in (L_u, L_v) \subseteq (l_1, \ldots, l_{2r}) \subseteq (f_1, \ldots, f_r)$$

Notice $R_1[X]$ is a free $(R[t_1, \ldots, t_p])[X]$-module with free basis given by all monic monomials in $u_1, \ldots, u_{2r}, v_1, \ldots, v_{2r}$. Therefore,

$$\begin{aligned}\mathcal{R}_X(L_u, L_v) &= \sum_{i=1}^{h} D_i(t_1, \ldots, t_p) M_i(u_1, \ldots, u_{2r}, v_1, \ldots, v_{2r}) \\ &\in (f_1, \ldots, f_r)\end{aligned}$$

implies $(D_1, \ldots, D_h) \subseteq (f_1, \ldots, f_r)$ in $(R[t_1, \ldots, t_p])[X]$. Thus, a resultant system of $f_1, \ldots, f_r$ is always contained in the ideal in $(R[t_1, \ldots, t_p])[X]$ generated by $f_1, \ldots, f_r$.

Now suppose K is a fixed field, and let $X_1, \ldots, X_n$ denote indeterminates over K. Consider r polynomials $f_1(X_1, \ldots, X_n), \ldots, f_r(X_1, \ldots, X_n)$ in

$K[X_1, \ldots, X_n]$. We can use Kronecker's method of elimination to devise an algorithm for finding a common solution (if any) to the following system of equations:

$$\textbf{C.9} \quad \begin{aligned} f_1(X_1, \ldots, X_n) &= 0 \\ &\vdots \\ f_r(X_1, \ldots, X_n) &= 0 \end{aligned}$$

If L is a field containing K, then $(z_1, \ldots, z_n)^t \in L^n$ is called a common solution to the equations in C.9 if $f_1(z_1, \ldots, z_n) = \cdots = f_r(z_1, \ldots, z_n) = 0$. Of course, the equations in C.9 may not have a common solution in L^n for any extension L of K. For example, if $A_1 f_1 + \cdots + A_r f_r = 1$ for some $A_1, \ldots, A_r \in K[X_1, \ldots, X_n]$, then clearly $f_1 = \cdots = f_r = 0$ have no common solution.

In order to study the common solutions of $f_1 = \cdots = f_r = 0$, we need a slightly different version of Theorem C.5. Suppose we write the polynomials appearing in equation C.9 as polynomials in X_n with coefficients in $K[X_1, \ldots, X_{n-1}]$.

$$\textbf{C.10} \quad \begin{aligned} f_1 &= a_0^{(1)} X_n^{\alpha(1)} + a_1^{(1)} X_n^{\alpha(1)-1} + \cdots + a_{\alpha(1)-1}^{(1)} X_n + a_{\alpha(1)}^{(1)} \\ &\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ f_r &= a_0^{(r)} X_n^{\alpha(r)} + a_1^{(r)} X_n^{\alpha(r)-1} + \cdots + a_{\alpha(r)-1}^{(r)} X_n + a_{\alpha(r)}^{(r)} \end{aligned}$$

As usual, we assume the formal degrees $\alpha(1), \ldots, \alpha(r)$ are all positive. The coefficients $\{a_j^{(i)} \mid 0 \leq j \leq \alpha(i);\ 1 \leq i \leq r\}$ are all polynomials in $K[X_1, \ldots, X_{n-1}]$. If we replace R with K in Theorem C.5, we obtain the following result:

Theorem C.11 With the notation as in equation C.10, there exist polynomials $D_1, \ldots, D_h \in K[X_1, \ldots, X_{n-1}]$ with the following property: Let L be a field containing K, and let $\sigma : K[X_1, \ldots, X_{n-1}] \mapsto L$ be a K-algebra homomorphism such that $\sigma(X_i) = z_i$ for $i = 1, \ldots, n-1$. Then $f_1^\sigma(X_n) = \cdots = f_r^\sigma(X_n) = 0$ have a common solution in some extension of L, or

$$a_0^{(1)}(z_1, \ldots, z_{n-1}) = \cdots = a_0^{(r)}(z_1, \ldots, z_{n-1}) = 0$$

if and only if

$$D_1(z_1, \ldots, z_{n-1}) = \cdots = D_h(z_1, \ldots, z_{n-1}) = 0 \qquad \blacksquare$$

We can use Theorem C.11 to devise an algorithm which will allow us to decide if the equations in C.9 have a common solution. We need to make one adjustment to the base field K. Suppose K' is a field containing K. Then we can view C.9 as a system of polynomial equations with coefficients in K'. Obviously, $f_1 = \cdots = f_r = 0$ have a common solution in L^n for some field $L \supseteq K$ if and only if $f_1 = \cdots = f_r = 0$ have a common solution in $(L')^n$ for some field $L' \supseteq K'$. In particular, by enlarging K if need be, we can always assume K is an infinite field when discussing the existence of a common solution of the equations in C.9. We will make this assumption throughout the rest of this discussion.

Theorem C.11 suggests that common solutions $D_1(X_1, \ldots, X_{n-1}) = \cdots = D_h(X_1, \ldots, X_{n-1}) = 0$ produce common solutions to $f_1(X_1, \ldots, X_n) = \cdots = f_r(X_1, \ldots, X_n)$ provided we can do away with the degenerate situation $a_0^{(1)}(z_1, \ldots, z_{n-1}) = \cdots = a_0^{(r)}(z_1, \ldots, z_{n-1}) = 0$. If the system of equations in C.9 is nontrivial, we will show there is a second system of equations $f_1' = \cdots = f_r' = 0$ with the following two properties:

(a) $f_1 = \cdots = f_r = 0$ have a common solution in L^n for some field $L \supseteq K$ if and only if $f_1' = \cdots = f_r' = 0$ have a common solution in L^n.

(b) The polynomials $f_1', \ldots, f_r'$ have a representation as in equation C.10 with the formal coefficient of f_1' a nonzero constant in K.

The system of equations appearing in C.9 is said to be trivial if $f_i(X_1, \ldots, X_n) \in K$ for all $i = 1, \ldots, r$. Thus, if every polynomial appearing in C.9 is a constant in K, then the system is called trivial. Suppose the system of equations $f_1 = \cdots = f_r = 0$ is trivial. If every f_i is the zero polynomial, then any vector in L^n, L a field containing K, is a common solution of $f_1 = \cdots = f_r = 0$. If some f_i is a nonzero constant, then $f_1 = \cdots = f_r = 0$ have no common solution in any L^n. This is the complete story for trivial systems of equations. Hence, in what follows, we will assume the system of equations appearing in C.9 is nontrivial. This means at least one polynomial $f_i(X_1, \ldots, X_n)$ appearing in C.9 has positive degree with respect to at least one of the variables $X_1, \ldots, X_n$.

Suppose the system of equations in C.9 is nontrivial. After permuting the f_i if need be, we can assume $f_1(X_1, \ldots, X_n)$ has positive degree with respect to at least one of the variables $X_1, \ldots, X_n$. Since $f_1(X_1, \ldots, X_n)$ is not a constant in K, f_1 can be written as a sum of homogeneous polynomials in the following way:

$$f_1(X_1, \ldots, X_n) = h_0(X_1, \ldots, X_n) + \cdots + h_\alpha(X_1, \ldots, X_n)$$

Here each $h_j(X_1, \ldots, X_n)$ is a homogeneous polynomial of degree j in $K[X_1, \ldots, X_n]$, α is a positive integer, and $h_\alpha(X_1, \ldots, X_n) \neq 0$. Since K is infinite, there exists a linear change of variables of the form:

C.12
$$\begin{aligned} X_1 &= Y_1 + c_1Y_n \\ &\cdot \quad\quad \cdot \\ &\cdot \quad\quad \cdot \\ &\cdot \quad\quad \cdot \\ X_{n-1} &= Y_{n-1} + c_{n-1}Y_n \quad (\text{here } c_i \in K^* \text{ for all } i = 1, \ldots, n) \\ X_n &= c_nY_n \end{aligned}$$

such that

$$f_1' = f_1(Y_1 + c_1Y_n, \ldots, Y_{n-1} + c_{n-1}Y_n, c_nY_n) = b_0Y_n^\alpha + b_1Y_n^{\alpha-1} + \cdots + b_{\alpha-1}Y_n + b_\alpha$$

with $b_0 \in K^*$. The rest of the coefficients $b_1, \ldots, b_\alpha$ here are polynomials in $K[Y_1, \ldots, Y_{n-1}]$. To see this, merely substitute the equations in C.12 into $h_\alpha(X_1, \ldots, X_n)$ and choose $c_1, \ldots, c_n$ accordingly. Since K is infinite, there is always a choice of constants $c_1, \ldots, c_n \in K^*$ (see the proof of Lemma 8.6 in the text) such that

$$f_1(Y_1 + c_1Y_n, \ldots, Y_{n-1} + c_{n-1}Y_n, c_nY_n) = b_0Y_n^\alpha + b_1Y_n^{\alpha-1} + \cdots + b_\alpha$$

has a nonzero constant $b_0 \in K$ as its leading coefficient of Y_n^α.

Set

$$f_i'(Y_1, \ldots, Y_n) = f_i(Y_1 + c_1Y_n, \ldots, Y_{n-1} + c_{n-1}Y_n, c_nY_n)$$

for each $i = 1, \ldots, r$. It is easy to see that $f_1 = \cdots = f_r = 0$ have a common solution in L^n for some extension field $L \supseteq K$ if and only if $f_1' = \cdots = f_r' = 0$ have a common solution in L^n. Since $b_0 \in K^*$, the polynomials $f_1', \ldots, f_r'$ have a representation (with respect to the variables $Y_1, \ldots, Y_n$) as in equation C.10 with the formal leading coefficient of f_1' being b_0.

Now let $D_1(Y_1, \ldots, Y_{n-1}), \ldots, D_h(Y_1, \ldots, Y_{n-1})$ be a resultant system of $f_1', \ldots, f_r'$ with respect to Y_n (Theorem C.11). Since $b_0 \in K^*$, $b_0(z_1, \ldots, z_{n-1}) = b_0 \neq 0$ for any $z_1, \ldots, z_{n-1} \in L \supseteq K$. It follows from Theorem C.11 that any common solution to $D_1 = \cdots = D_h = 0$ produces a common solution to $f_1' = \cdots = f_r' = 0$. If $D_1 = \cdots = D_h = 0$ have no common solution, then $f_1' = \cdots = f_r' = 0$ have no common solution. Thus, the system $f_1 = \cdots = f_r = 0$ in n variables $X_1, \ldots, X_n$ has been replaced by the system $D_1 = \cdots = D_h = 0$ in $n - 1$ variables $Y_1, \ldots, Y_{n-1}$.

We can now repeat this entire procedure on the equations $D_1(Y_1, \ldots, Y_{n-1}) = \cdots = D_h(Y_1, \ldots, Y_{n-1}) = 0$. If this system of equations is trivial, then Theorem C.11 implies $f_1 = \cdots = f_r = 0$ have a common solution if and only if $D_1, \ldots, D_h$ are all the zero polynomial. If the system $D_1 = \cdots = D_h =$

0 is nontrivial, eliminate another variable. Since the number of variables is finite, this procedure will decide in a finite number of steps if $f_1 = \cdots = f_r = 0$ have a common solution.

The algorithm described above is commonly called elimination theory. Theoretically, at least, we can always tell in a finite number of steps whether a given set of polynomial equations has a common solution in some extension field. The algorithm is difficult to apply in practice since the resultants needed are often very large. However, there is one corollary of this algorithm which has important consequences in algebraic geometry.

Theorem C.13 Let $f_1, \ldots, f_r \in K[X_1, \ldots, X_n]$. The equations $f_1 = \cdots = f_r = 0$ have no common solution in L^n for any algebraic extension L of K if and only if $A_1 f_1 + \cdots + A_r f_r = 1$ for some $A_1, \ldots, A_r \in K[X_1, \ldots, X_n]$.

Proof. Suppose $\sum_{i=1}^{r} A_i f_i = 1$ for some $A_1, \ldots, A_r \in K[X_1, \ldots, X_n]$. If $(z_1, \ldots, z_n)^t \in L^n$ is a common solution to $f_1 = \cdots = f_r = 0$, then in L,

$$0 = \sum_{i=1}^{r} A_i(z_1, \ldots, z_n) f_i(z_1, \ldots, z_n) = 1$$

This is clearly impossible. Hence, $f_1 = \cdots = f_r = 0$ have no common solution in L^n for any field $L \supseteq K$.

Suppose $f_1 = \cdots = f_r = 0$ have no common solution in L^n for any algebraic extension L of K. We proceed by induction on n. If $n = 1$, then the polynomials $f_1(X), \ldots, f_r(X)$ have no common root in $\overline{K}$, an algebraic closure of K. This implies the polynomials $f_1(X), \ldots, f_r(X)$ are relatively prime in $\overline{K}[X]$. But then, the polynomials $f_1(X), \ldots, f_r(X)$ are relatively prime in $K[X]$. Thus, there exist $A_1, \ldots, A_r \in K[X]$ such that $\sum_{i=1}^{r} A_i(X) f_i(X) = 1$.

Suppose we have established the result for any field K and any polynomials in $n - 1$ variables ($n \geq 2$) over K. Let $f_1, \ldots, f_r \in K[X_1, \ldots, X_n]$. Suppose $f_1 = \cdots = f_r = 0$ have no common solution in L^n for any algebraic extension L of K. We can assume that the system of equations $f_1 = \cdots = f_r = 0$ is nontrivial and that $f_1(X_1, \ldots, X_n)$ is not a constant in K.

Let $\overline{K}$ be an algebraic closure of K. We can view $f_1, \ldots, f_r$ as polynomials in $\overline{K}[X_1, \ldots, X_n]$. The field $\overline{K}$ is infinite. In particular, we can change variables as in equation C.12. Define

$$f_i'(Y_1, \ldots, Y_n) = f_i(Y_1 + c_1 Y_n, \ldots, Y_{n-1} + c_{n-1} Y_n, c_n Y_n) \in \overline{K}[Y_1, \ldots, Y_n]$$

for $i = 1, \ldots, r$. The constants $c_1, \ldots, c_n \in (\overline{K})^*$ are chosen such that

$$f_1'(Y_1, \ldots, Y_n) = b_0^{(1)} Y_n^{\alpha(1)} + b_1^{(1)} Y_n^{\alpha(1)-1} + \cdots + b_{\alpha(1)}^{(1)}$$

with $b_0^{(1)} \in (\overline{K})^*$. The system $f_1 = \cdots = f_r = 0$ has a common solution in $(\overline{K})^n$ if and only if the system $f_1' = \cdots = f_r' = 0$ has a common solution in $(\overline{K})^n$. Since $f_1 = \cdots = f_r = 0$ have no common solution in L^n for any algebraic extension L of K, $f_1' = \cdots = f_r' = 0$ have no common solution in $(\overline{K})^n$. Let $D_1, \ldots, D_h \in \overline{K}[Y_1, \ldots, Y_{n-1}]$ be a resultant system of $f_1', \ldots, f_r'$ with respect to Y_n. Theorem C.11 implies $D_1 = \cdots = D_h = 0$ have no common solution in $(\overline{K})^{n-1}$. Our induction hypothesis (applied to $\overline{K}$, and $D_1, \ldots, D_h$) implies $C_1D_1 + \cdots + C_hD_h = 1$ for some $C_1, \ldots, C_h \in \overline{K}[Y_1, \ldots, Y_{n-1}]$.

We have noted before that the ideal generated by $D_1, \ldots, D_h$ is contained in the ideal generated by $f_1', \ldots, f_r'$ in $\overline{K}[Y_1, \ldots, Y_n]$. Thus, $B_1f_1' + \cdots + B_rf_r' = 1$ for some $B_1, \ldots, B_r \in \overline{K}[Y_1, \ldots, Y_n]$. Since $c_n \in (\overline{K})^*$,

$$Y_1 = X_1 - c_1c_n^{-1}X_n, \ldots, Y_{n-1} = X_{n-1} - c_{n-1}c_n^{-1}X_n, Y_n = c_n^{-1}X_n$$

Therefore,

$$\begin{aligned} 1 &= \sum_{i=1}^{r} B_i(Y_1, \ldots, Y_n)f_i'(Y_1, \ldots, Y_n) \\ &= \sum_{i=1}^{r} B_i(Y_1, \ldots, Y_n)f_i(Y_1 + c_1Y_n, \ldots, Y_{n-1} + c_{n-1}Y_n, c_nY_n) \\ &= \sum_{i=1}^{r} B_i(X_1 - c_1c_n^{-1}X_n, \ldots, X_{n-1} - c_{n-1}c_n^{-1}X_n, c_n^{-1}X_n)f_i(X_1, \ldots, X_n) \end{aligned}$$

Therefore, $A_1'f_1 + \cdots + A_r'f_r = 1$ for some $A_1', \ldots, A_r' \in \overline{K}[X_1, \ldots, X_n]$.

Since K is a field, $\overline{K}$ is a free K-module. It easily follows from this that $\overline{K}[X_1, \ldots, X_n]$ is a free $K[X_1, \ldots, X_n]$-module. We can always find a free $K[X_1, \ldots, X_n]$-module basis $\{z_j \mid j \in \Gamma\}$ of $\overline{K}[X_1, \ldots, X_n]$ such that some z_j is 1 (use Zorn's lemma). Writing each A_i' as a linear combination of the z_j with coefficients in $K[X_1, \ldots, X_n]$ and substituting in $A_1'f_1 + \cdots + A_r'f_r = 1$, we have $A_1f_1 + \cdots + A_rf_r = 1$ for some $A_1, \ldots, A_r \in K[X_1, \ldots, X_n]$. ∎

The following corollary to Theorem C.13 is usually called Hilbert's Nullstellensatz.

Corollary C.14 Let $f_1, \ldots, f_r$ and $f \in K[X_1, \ldots, X_n]$. Suppose $f(\xi_1, \ldots, \xi_n) = 0$ for all common zeros $(\xi_1, \ldots, \xi_n)^t$ of $f_1 = \cdots = f_r = 0$. Then there exist a positive integer p and polynomials $A_1, \ldots, A_r \in K[X_1, \ldots, X_n]$ such that $f^p = A_1f_1 + \cdots + A_rf_r$.

Thus, if f vanishes at the common zeros of $f_1, \ldots, f_r$, then $f \in \sqrt{(f_1, \ldots, f_r)}$.

Proof. The result is certainly true if $f = 0$. Hence, we may assume $f \neq 0$. Let Z denote an indeterminate over $K[X_1, \ldots, X_n]$, and consider the polynomials $f_1, \ldots, f_r, f_{r+1} = 1 - Zf$ in $K[X_1, \ldots, X_n, Z]$. Since f vanishes at every common zero of $f_1, \ldots, f_r$, the equations $f_1 = \cdots = f_r = f_{r+1} = 0$ have no common solution in L^{n+1} for any algebraic extension L of K. Theorem C.13 implies $C_1 f_1 + \cdots + C_r f_r + C f_{r+1} = 1$ for some $C, C_1, \ldots, C_r \in K[X_1, \ldots, X_n, Z]$. Since Z is an indeterminate over $K[X_1, \ldots, X_n]$, we can substitute $1/f$ for Z in this equation. Thus, $1 = \sum_{i=1}^{r} C_i(X_1, \ldots, X_n, 1/f) f_i(X_1, \ldots, X_n)$. Clearing denominators by multiplying by a suitable power of f^p gives the result. ∎

The reader is urged to consult various texts on algebraic geometry to see how these results are used. For example, [S] is a good introduction to algebraic geometry. Our second application of resultants is taken from this text. One of the most famous results concerning plane curves is the following theorem of Bezout:

Theorem C.15 (Bezout's Theorem) Let Γ and Δ be two plane curves in $\mathbb{P}_K^2$ of orders n and m, respectively. If Γ and Δ have no common components, then Γ and Δ intersect in mn points counting their multiplicities.

In order that the reader may understand what is being said here as well as making clear what connections this theorem has with the theory of resultants, we will give a brief description of all the terms appearing in Theorem C.15. For more details, the reader can consult [S].

Throughout this discussion, K will denote an algebraically closed field of arbitrary characteristic. The set $\mathbb{P}_K^n$ is called projective n-space over K. It is constructed from $K^{n+1} - \{(0, \ldots, 0)^t\}$ in the following way: Define an equivalence relation $\equiv$ on the set $P = K^{n+1} - \{(0, \ldots, 0)^t\}$ by setting $(x_1, \ldots, x_{n+1})^t \equiv (x_1', \ldots, x_{n+1}')^t$ if there exists an element $c \in K^*$ such that $cx_i = x_i'$ for all $i = 1, \ldots, n + 1$. It is easy to check that the relation $\equiv$ is reflexive, symmetric, and transitive. Thus, $\equiv$ is an equivalence relation on P. The set of equivalence classes of $(P, \equiv)$ will be denoted by $\mathbb{P}_K^n$. (See [B, Section 5, Chapter I] for more information on equivalence relations and classes.) If $(x_1, \ldots, x_{n+1})^t \in P$, then the equivalence class containing $(x_1, \ldots, x_{n+1})^t$ will be denoted by $(x_1 : \cdots : x_{n+1})$. Thus,

$$(x_1 : \cdots : x_{n+1}) = \{(z_1, \ldots, z_{n+1})^t \in K^{n+1} - \{(0, \ldots, 0)^t\} \mid \text{there exists an element } c \in K^* \text{ such that } cz_i = x_i \text{ for all } i = 1, \ldots, n + 1\}$$

The set $(x_1 : \cdots : x_{n+1})$ is called a point in $\mathbb{P}_K^n$, and

$$\mathbb{P}_K^n = \{(x_1 : \cdots : x_{n+1}) \mid (x_1, \ldots, x_{n+1})^t \in K^{n+1} - \{(0, \ldots, 0)^t\}\}$$

If $(x_1 : \cdots : x_{n+1})$ is a point in $\mathbb{P}_K^n$, then the elements $x_1, \ldots, x_{n+1}$ are called coordinates of $(x_1 : \cdots : x_{n+1})$. In particular, $y_1, \ldots, y_{n+1}$ and $x_1, \ldots, x_{n+1}$ are coordinates of the same point in $\mathbb{P}_K^n$ if and only if there exists an element $c \in K^*$ such that $cy_i = x_i$ for all $i = 1, \ldots, n+1$. Notice that any point of $\mathbb{P}_K^n$ has at least one coordinate which is nonzero. We will be interested in the sets $\mathbb{P}_K^1$ and $\mathbb{P}_K^2$. The set $\mathbb{P}_K^1$ is called the projective line. $\mathbb{P}_K^2$ is called the projective plane.

Let $R = K[X_1,X_2,X_3]$. Suppose $f(X_1,X_2,X_3)$ is a nonconstant, homogeneous polynomial in R of degree d. Then f can be written in the following way:

$$\textbf{C.16} \quad f(X_1,X_2,X_3) = \sum_{\alpha(1)+\alpha(2)+\alpha(3)=d} c_{\alpha(1)\alpha(2)\alpha(3)} X_1^{\alpha(1)} X_2^{\alpha(2)} X_3^{\alpha(3)}$$

In equation C.16, the sum is taken over all nonnegative integers $\alpha(1)$, $\alpha(2)$, and $\alpha(3)$ whose sum is d, the degree of f. Since f is not a constant, $d \geq 1$. The coefficients $c_{\alpha(1)\alpha(2)\alpha(3)}$ are elements of K. At least one $c_{\alpha(1)\alpha(2)\alpha(3)}$ is nonzero since $f \neq 0$. Let $a,b,c \in K$. Since f is homogeneous, $f(a,b,c) = 0$ if and only if $f(ta,tb,tc) = 0$ for all $t \in K^*$. In particular, the set

$$\textbf{C.17} \quad V(f) = \{(a : b : c) \in \mathbb{P}_K^2 \mid f(a,b,c) = 0\}$$

is a well-defined subset of $\mathbb{P}_K^2$. Since K is an algebraically closed field, $V(f)$ contains infinitely many distinct points of P_K^2. Also, it is easy to see $V(f)$ is a proper subset of $\mathbb{P}_K^2$. The set $V(f)$ is called an algebraic (plane) curve defined by the equation $f(X_1,X_2,X_3) = 0$.

Suppose $\Gamma = V(f)$ is an algebraic curve in $\mathbb{P}_K^2$. The curve Γ is said to be reducible if Γ is the proper union of two other curves Γ_1 and Γ_2. Thus, Γ is reducible if there exist algebraic curves $\Gamma_1, \Gamma_2 \subseteq P_K^2$ such that $\Gamma = \Gamma_1 \cup \Gamma_2$ and $\Gamma_1 < \Gamma, \Gamma_2 < \Gamma$. If Γ is not reducible, then Γ is called an irreducible curve. Since R is a unique factorization domain, f can be factored in R as a product of irreducible polynomials: $f = f_1^{r(1)} f_2^{r(2)} \cdots f_p^{r(p)}$. Here, $r(1), \ldots, r(p)$ are positive integers, and $f_1, \ldots, f_p$ are homogeneous, irreducible polynomials which are pairwise nonassociates. It is easy to check that $V(f) = V(f_1) \cup \cdots \cup V(f_p)$, and each $V(f_i)$ is an irreducible curve in $\mathbb{P}_K^2$. The decomposition $V(f) = V(f_1) \cup \cdots \cup V(f_p)$ into irreducible curves is unique. The curves $V(f_1), \ldots, V(f_p)$ are called the (irreducible) components of $V(f)$. Two curves $\Gamma, \Delta \subseteq \mathbb{P}_K^2$ have no common components if no irreducible component of Γ is an irreducible component of Δ and vice versa. For example, $\Gamma = V(X_1)$ and $\Delta = V(X_2)$ are two irreducible curves in $\mathbb{P}_K^2$ and, consequently, have no common components. If Γ and Δ are two curves in $\mathbb{P}_K^2$, then Γ and Δ intersect in at most finitely many points of $\mathbb{P}_K^2$ if and only if Γ and Δ have no common components. A proof of this fact can be found in [S, Theorem 3.14].

The order of a curve Γ is the degree of the (homogeneous) polynomial f of smallest degree for which $V(f) = \Gamma$. We will let order(Γ) denote the order of Γ. Thus, order(Γ) = min$\{\partial(f) \mid f$ is homogeneous, and $V(f) = \Gamma\}$. If $\Gamma = V(f_1) \cup \cdots \cup V(f_p)$ is a decomposition of Γ into irreducible components with each f_i irreducible, then clearly order(Γ) = $\sum_{i=1}^{p} \partial(f_i)$.

We have now defined all of the terms, $\mathbb{P}_K^2$, curves, order, and components, appearing in Bezout's theorem. The last term, multiplicities, is defined using resultants. As in elimination theory, it is convenient to make certain changes of variables. These changes are called homogeneous, linear transformations of $\mathbb{P}_K^2$.

Suppose $A = (a_{ij}) \in \mathrm{Gl}(3,K)$. The nonsingular matrix A induces a map $L_A : \mathbb{P}_K^2 \mapsto \mathbb{P}_K^2$ defined by the following equation:

$$L_A((x_1 : x_2 : x_3)) = \left(\sum_{j=1}^{3} a_{1j}x_j : \sum_{j=1}^{3} a_{2j}x_j : \sum_{j=1}^{3} a_{3j}x_j\right)$$

Thus, $L_A((x_1 : x_2 : x_3))$ is the equivalence class of the ordered triple $A(x_1,x_2,x_3)^t$. It is easy to see that L_A is a well-defined function which maps $\mathbb{P}_K^2$ bijectively onto itself. The function L_A is called a homogeneous, linear transformation of $\mathbb{P}_K^2$.

Suppose $L_A : \mathbb{P}_K^2 \mapsto \mathbb{P}_K^2$ is a homogeneous, linear transformation. Let $\mathfrak{H}$ denote the set of all homogeneous polynomials in R. For this discussion, it is convenient to regard 0 as a homogeneous polynomial of any degree. Associated with L_A is the K-automorphism $\tau_A : R \mapsto R$ given by

$$\tau_A(F(X_1,X_2,X_3)) = F\left(\sum_{j=1}^{3} b_{1j}X_j, \sum_{j=1}^{3} b_{2j}X_j, \sum_{j=1}^{3} b_{3j}X_j\right)$$

where $(b_{ij}) = A^{-1}$. Clearly, $\tau_A(\mathfrak{H}) = \mathfrak{H}$.

There are several rules concerning the relationships between L_A and τ_A. We list these in C.18.

C.18 (a) If f is homogeneous of degree d, so is $\tau_A(f)$.
(b) $f = 0$ if and only if $\tau_A(f) = 0$.
(c) $L_A(V(f)) = V(\tau_A(f))$ for any $f \in \mathfrak{H}$.
(d) f is irreducible if and only if $\tau_A(f)$ is irreducible.
(e) f and g are associates in R if and only if $\tau_A(f)$ and $\tau_A(g)$ are associates in R.
(f) An equation of minimal degree for a curve Γ is mapped by τ_A into an equation of minimal degree for $L_A(\Gamma)$. Thus, order(Γ) = order($L_A(\Gamma)$) for any algebraic curve $\Gamma \subseteq \mathbb{P}_K^2$.

(g) Let P_1,P_2,P_3,P_4 be four points in $\mathbb{P}_K^2$ no three of which are collinear. Let Q_1,Q_2,Q_3,Q_4 be another four points of $\mathbb{P}_K^2$ no three of which are collinear. Then there exists a unique $L_A : \mathbb{P}_K^2 \mapsto \mathbb{P}_K^2$ such that $L_A(P_i) = Q_i$ for $i = 1, \ldots, 4$.

(h) If Γ and Δ are algebraic curves in $\mathbb{P}_K^2$ which intersect in at most finitely many points, there exists a homogeneous, linear transformation $L_A : \mathbb{P}_K^2 \mapsto \mathbb{P}_K^2$ such that $\Gamma' = L_A(\Gamma)$ and $\Delta' = L_A(\Delta)$ have the following properties:
   (i) $(0:0:1) \notin \Gamma'$, or $(0:0:1) \notin \Delta'$
   (ii) No pair of intersection points of Γ' and Δ' are collinear with $(0:0:1)$.

In C.18, points are said to be collinear if their coordinates satisfy the equation of a line: $aX_1 + bX_2 + cX_3 = 0$ in $\mathbb{P}_K^2$. The assertions in C.18 are all easy to prove. (See [S] for more details.) Two curves which satisfy conditions (i) and (ii) in C.18h are said to be in permissible position. The assertion in C.18h is that two curves without common components can always be transformed by some homogeneous, linear transformation L_A into curves in permissible position.

We can now use the theory of resultants to count the number of intersections of two curves without common components.

Theorem C.19 Let Γ and Δ be two algebraic curves in $\mathbb{P}_K^2$ of orders n and m, respectively. Suppose Γ and Δ have no common components. Then Γ and Δ intersect in at least one point and at most mn points.

Proof. Since the order of Γ is n, $\Gamma = V(f)$ for some homogeneous polynomial $f \in R^*$ with $\partial(f) = n$. Similarly, $\Delta = V(g)$ for some homogeneous polynomial $g \in R^*$ with $\partial(g) = m$. Since Γ and Δ have no common components, we have already noted that $\Gamma \cap \Delta$ consists of at most a finite number of points in $\mathbb{P}_K^2$. By C.18h, there exists a homogeneous, linear transformation $L_A : \mathbb{P}_K^2 \mapsto \mathbb{P}_K^2$ such that $\Gamma' = L_A(\Gamma)$ and $\Delta' = L_A(\Delta)$ are in permissible position. $\Gamma' = V(\tau_A(f))$, and $\Delta' = V(\tau_A(g))$ with $\partial(\tau_A(f)) = \partial(f) = n$ and $\partial(\tau_A(g)) = \partial(g) = m$. $P_1, \ldots, P_s$ are points in $\Gamma \cap \Delta$ if and only if $L_A(P_1), \ldots, L_A(P_s)$ are points in $\Gamma' \cap \Delta'$. Hence, it suffices to prove the theorem for Γ' and Δ'. In other words, we can assume without loss of generality that Γ and Δ are in permissible position in $\mathbb{P}_K^2$.

Write f and g as polynomials in X_3 with coefficients in $K[X_1,X_2]$.

$$\text{C.20} \quad \begin{aligned} f(X_1,X_2,X_3) &= a_0X_3^n + a_1(X_1,X_2)X_3^{n-1} + \cdots + a_n(X_1,X_2) \\ g(X_1,X_2,X_3) &= b_0X_3^m + b_1(X_1,X_2)X_3^{m-1} + \cdots + b_m(X_1,X_2) \end{aligned}$$

Each $a_i(X_1,X_2)$ [or $b_i(X_1,X_2)$] is a homogeneous polynomial in $K[X_1,X_2]$ of degree i, for $i = 0, \ldots, n[m]$. Since Γ and Δ are in permissible position,

$(0:0:1) \notin \Gamma$ or $(0:0:1) \notin \Delta$. We may assume $(0:0:1) \notin \Gamma$. Then a_0 is a nonzero constant in K. On the other hand, the constant b_0 could be zero.

Set

$$H(X_1,X_2) = \mathcal{R}_{X_3}(f,g) = S(a_0, \ldots, a_n, b_0, \ldots, b_m)$$

Theorem 8.28 implies $H(X_1,X_2)$ is either zero or a nonzero, homogeneous polynomial of degree mn in $K[X_1,X_2]$. Suppose $H(X_1,X_2)$ is the zero polynomial. Then for every $z_1,z_2 \in K$,

$$0 = H(z_1,z_2) = \mathcal{R}_{X_3}(f(z_1,z_2,X_3), g(z_1,z_2,X_3))$$

The theorem in C.2 then implies $f(z_1,z_2,X_3) = g(z_1,z_2,X_3) = 0$ have a common solution $z_3 \in K$. But then Γ and Δ have infinitely many points in common. This is contrary to our assumption that Γ and Δ have no common components. We conclude that $H(X_1,X_2)$ is a nonzero, homogeneous polynomial of degree mn in $K[X_1,X_2]$.

We will prove the theorem by showing that the points in $\Gamma \cap \Delta$ are in a one-to-one correspondence with the solutions of $H(X_1,X_2) = 0$ on the projective line $\mathbb{P}^1_K$. Suppose $(a:b:c) \in \Gamma \cap \Delta$. Then either $a \neq 0$ or $b \neq 0$. For otherwise, $(0:0:1) \in \Gamma$. In particular, $(a:b)$ is a well-defined point in $\mathbb{P}^1_K$. Since $(a:b:c) \in \Gamma \cap \Delta, f(a,b,c) = g(a,b,c) = 0$. Therefore, $f(a,b,X_3)$ and $g(a,b,X_3)$ have a common factor $X_3 - c$ in $K[X_3]$. Theorem C.2 implies $H(a,b) = \mathcal{R}_{X_3}(f(a,b,X_3), g(a,b,X_3)) = 0$. Thus, the point $(a:b)$ is a solution of the equation $H(X_1,X_2) = 0$ in $\mathbb{P}^1_K$. Define a map $\vartheta : \Gamma \cap \Delta \mapsto \{(x_1:x_2) \in \mathbb{P}^1_K \mid H(x_1,x_2) = 0\}$ by setting $\vartheta((a:b:c)) = (a:b)$. The above discussion shows ϑ is a well defined function.

Since K is an algebraically closed field and H is a homogeneous polynomial of degree mn, $H(X_1,X_2)$ can be factored in $K[X_1,X_2]$ as follows:

$$\textbf{C.21} \quad H(X_1,X_2) = \prod_{i=1}^{mn} (y_1^{(i)}X_2 - y_2^{(i)}X_1).$$

In equation C.21, $y_1^{(1)}, \ldots, y_1^{(mn)}$ and $y_2^{(1)}, \ldots, y_2^{(mn)}$ are constants in K. Since $H(X_1,X_2) \neq 0$, each pair $(y_1^{(i)},y_2^{(i)})$ cannot be identically zero. Therefore, $\Lambda = \{(y_1^{(i)} : y_2^{(i)}) \mid i = 1, \ldots, mn\}$ is a finite set of points in $\mathbb{P}^1_K$. The pairs $(y_1^{(i)},y_2^{(i)})$ are not necessarily distinct. Hence, Λ contains at most mn distinct points of $\mathbb{P}^1_K$. It follows readily from equation C.21 that Λ is the complete set of solutions to $H(X_1,X_2) = 0$ in $\mathbb{P}^1_K$. Thus, $\Lambda = \{(a:b) \in \mathbb{P}^1_K \mid H(a,b) = 0\}$. In particular, $\operatorname{Im} \vartheta \subseteq \Lambda$.

Any point $(a:b) \in \Lambda$ determines a point $(a:b:c) \in \Gamma \cap \Delta$ by C.2. Thus, the map $\vartheta : \Gamma \cap \Delta \mapsto \Lambda$ is surjective. If $(a:b:c)$ and $(a:b:c')$ are both points of $\Gamma \cap \Delta$, then $(a:b:c)$, $(a:b:c')$, and $(0:0:1)$ all lie on the line $aX_2 - bX_1 = 0$. Since Γ and Δ are in permissible position, we conclude $c =$

c'. Thus, $\vartheta^{-1}\{(a : b)\}$ is a single point in $\Gamma \cap \Delta$. In particular, ϑ is injective, and the points of $\Gamma \cap \Delta$ are in a one-to-one correspondence with the points in Λ. Therefore, $\Gamma \cap \Delta$ contains at most mn points. Since $\Lambda \neq \emptyset$, $\Gamma \cap \Delta$ contains at least one point. This completes the proof of Theorem C.19. ∎

The proof of Theorem C.19 contains an algorithm for determining the intersection points of two curves Γ and Δ which have no common components. First, find homogeneous polynomials f and g of minimal degrees such that $\Gamma = V(f)$ and $\Delta = V(g)$. Second, apply a suitable homogeneous, linear transformation L_A to $\mathbb{P}_K^2$ such that $\Gamma' = L_A(\Gamma)$ and $\Delta' = L_A(\Delta)$ are in permissible position. The equations for Γ' and Δ' are $\tau_A(f) = f' = 0$ and $\tau_A(g) = g' = 0$, respectively. Third, compute $\mathcal{R}_{X_3}(f',g')$, and factor this polynomial as in equation C.21. Fourth, for each $(y_1^{(i)},y_2^{(i)})$ occurring in the factorization of $\mathcal{R}_{X_3}(f',g')$, compute the common solution c_i to $f'(y_1^{(i)},y_2^{(i)},X_3) = g'(y_1^{(i)},y_2^{(i)},X_3) = 0$. Finally, pull the points $(y_1^{(i)} : y_2^{(i)} : c_i)$ back to $\Gamma \cap \Delta$ via L_A^{-1}. This procedure is obviously a special case of Kronecker's method of elimination.

Suppose Γ and Δ are two algebraic curves in $\mathbb{P}_K^2$ having no common components. Suppose order$(\Gamma) = n$ and order$(\Delta) = m$. Then $\Gamma = V(f)$ for some homogeneous polynomial $f \in R$ of degree n and $\Delta = V(g)$ for some homogeneous polynomial $g \in R$ of degree m. Set $\Gamma \cap \Delta = \{P_1, \ldots, P_s\}$. Then Theorem C.19 implies $1 \leq s \leq mn$. It is possible to define an integer $i(\Gamma,\Delta,P_j)$, called the intersection multiplicity of Γ and Δ at P_j, such that $\sum_{j=1}^{s} i(\Gamma,\Delta,P_j) = mn$.

If Γ and Δ are in permissible position in $\mathbb{P}_K^2$, then the proof of Theorem C.19 provides a ready definition of $i(\Gamma,\Delta,P_j)$. If $P_j = (y_1^{(j)} : y_2^{(j)} : c_j)$ for $j = 1, \ldots, s$, then

$$\textbf{C.22} \quad H(X_1,X_2) = \mathcal{R}_{X_3}(f,g) = \prod_{j=1}^{s} (y_1^{(j)}X_2 - y_2^{(j)}X_1)^{s(j)}$$

In equation C.22, $s(j)$ is the multiplicity of the linear factor $y_1^{(j)}X_2 - y_2^{(j)}X_1$ in $H(X_1,X_2)$. Define $i(\Gamma,\Delta,P_j) = s(j)$ for all $j = 1, \ldots, s$. Since H is a homogeneous polynomial of degree mn, we have $\sum_{j=1}^{s} i(\Gamma,\Delta,P_j) = \sum_{j=1}^{s} s(j) = mn$. The statement "$\Gamma$ and Δ intersect in mn points counting their multiplicities" in Theorem C.15 means precisely $\sum_{j=1}^{s} i(\Gamma,\Delta,P_j) = mn$. Thus, we have proved Bezout's theorem when Γ and Δ are in permissible position in $\mathbb{P}_K^2$.

When Γ and Δ are not in permissible position, the intersection multiplicity $i(\Gamma,\Delta,P_j)$ is defined as follows:

$$i(\Gamma,\Delta,P_j) = i(L_A(\Gamma), L_A(\Delta), L_A(P_j))$$

where $L_A : \mathbb{P}_K^2 \mapsto \mathbb{P}_K^2$ is a homogeneous, linear transformation taking Γ and Δ to curves $L_A(\Gamma)$ and $L_A(\Delta)$ in permissible position in $\mathbb{P}_K^2$. For this definition to make sense, we must argue that

$$i(L_A(\Gamma), L_A(\Delta), L_A(P_j)) = i(L_B(\Gamma), L_B(\Delta), L_B(P_j))$$

for any two homogeneous, linear transformations L_A and L_B which take Γ and Δ to curves in permissible position. A proof of this fact can be found in [S, Chapter 7]. Thus, we have the following more precise version of Theorem C.15:

Theorem C.23 (Bezout's Theorem) Let Γ and Δ be two algebraic curves of orders n and m respectively in $\mathbb{P}_K^2$. Suppose Γ and Δ have no common components. Then Γ and Δ intersect in a finite number of points $P_1, \ldots, P_s$, and $\sum_{j=1}^{s} i(\Gamma,\Delta,P_j) = mn$. ∎

There are many generalizations of Bezout's theorem to higher dimensions. For instance, we have the following theorem:

Theorem C.24 Suppose $f_1(X_1, \ldots, X_n), \ldots, f_{n-1}(X_1, \ldots, X_n)$ are $n-1$ homogeneous polynomials in $K[X_1, \ldots, X_n]$ such that the equations $f_1 = \cdots = f_{n-1} = 0$ have only finitely many common solutions in $\mathbb{P}_K^{n-1}$. Then the number of solutions to these equations counted with appropriate multiplicities is equal to the product of the degrees of $f_1, \ldots, f_{n-1}$.

The important part of this theorem is to define carefully what is meant by the multiplicity of a common solution of the equations $f_1 = \cdots = f_{n-1} = 0$ in $\mathbb{P}_K^{n-1}$. As above, the multiplicity of a common solution is computed from a certain resultant system derived from $f_1, \ldots, f_{n-1}$. An explanation of how this is done, as well as a proof of Theorem C.24, can be found in [W].

REFERENCES

[B] W. C. Brown, *A Second Course in Linear Algebra*, John Wiley & Sons, New York, 1988.

[J] N. Jacobson, *Basic Algebra I*, second edition, W. H. Freeman, New York, 1985.

[S] A. Seidenberg, *Elements of the Theory of Algebraic Curves*, Addison-Wesley, Reading, Massachusetts, 1968.

[W] B. L. Van Der Waerden, *Modern Algebra*, Vol. II, Frederick Unger, New York, 1940.

Appendix D
The Hilbert-Burch Theorem

In this appendix, we will use some of the ideas from Chapters 5 and 13 to prove the Hilbert-Burch theorem. Our proof of this result is derived from Geramita and Kani's treatment of this subject in [G]. Throughout this appendix, R will denote a Noetherian (commutative) ring.

Definition D.1 Let $x_1, \ldots, x_r \in R$. The r-tuple $(x_1, \ldots, x_r)^t \in R^r$ is called a regular sequence of R if the following two conditions are satisfied:

(a) $(x_1, \ldots, x_r) \neq R$.

(b) For each $i = 1, \ldots, r$, $x_i \notin Z(R/(x_1, \ldots, x_{i-1}))$.

In the above definition, $(x_1, \ldots, x_r)$ denotes the ideal of R generated by $x_1, \ldots, x_r$. Thus, $(x_1, \ldots, x_r) = \sum_{i=1}^{r} Rx_i$. If $(x_1, \ldots, x_r)^t$ is a regular sequence of R, then the ideal generated by the entries in the vector $(x_1, \ldots, x_r)^t$ is a proper ideal of R. When $i = 1$, condition (b) means x_1 is a regular element in the ring R. If $(x_1, \ldots, x_r)^t$ is a regular sequence, then x_1 is a regular element of R, the image of x_2 is a regular element in the ring $R/(x_1)$, the image of x_3 is a regular element in $R/(x_1,x_2)$, etc. Notice that the order in which the elements appear in the vector $(x_1, \ldots, x_r)^t$ is important here. It is possible for $(x_1, \ldots, x_r)^t$ to be a regular sequence of R while some permutation $(x_{\sigma(1)}, \ldots, x_{\sigma(r)})^t$ of $(x_1, \ldots, x_r)^t$ is not a regular sequence of R.

The simplest example of a regular sequence is provided by indeterminates.

Suppose $X_1, \ldots, X_n$ are indeterminates over a field F. Then $(X_1, \ldots, X_r)^t \in (F[X_1, \ldots, X_n])^r$ is a regular sequence of $R = F[X_1, \ldots, X_n]$ for each $r = 1, \ldots, n$.

If $\xi = (x_1, \ldots, x_r)^t \in R^r$ is a regular sequence of R, then the integer r is called the length of ξ. Let $\mathfrak{A}$ be a proper ideal of R, and suppose $\xi = (x_1, \ldots, x_r)^t \in R^r$ is a regular sequence of R. ξ is called a regular sequence of $\mathfrak{A}$ if $x_i \in \mathfrak{A}$ for all $i = 1, \ldots, r$. ξ is called a maximal, regular sequence of $\mathfrak{A}$ if there is no element $y \in \mathfrak{A}$ such that $(x_1, \ldots, x_r, y)^t$ is a regular sequence of $\mathfrak{A}$. Thus, $\xi = (x_1, \ldots, x_r)^t \in R^r$ is a maximal, regular sequence of $\mathfrak{A}$ if ξ is a regular sequence of R, $x_i \in \mathfrak{A}$ for all $i = 1, \ldots, r$, and $\mathfrak{A} \subseteq Z(R/(x_1, \ldots, x_r))$. For example, the reader can easily check that $\xi = (X_1, \ldots, X_n)^t$ is a maximal, regular sequence of the ideal $\mathfrak{M} = (X_1, \ldots, X_n) \subseteq R = F[X_1, \ldots, X_n]$.

Again let $\mathfrak{A}$ be a proper ideal of R. Since R is Noetherian, maximal, regular sequences of $\mathfrak{A}$ exist. Furthermore, any two maximal, regular sequences of $\mathfrak{A}$ have the same length. Proofs of these two assertions can be found in [K, Chapter III]. In particular, the following definition makes sense.

Definition D.2 Let $\mathfrak{A}$ be a proper ideal of a Noetherian ring R. The grade of $\mathfrak{A}$ is the length of any maximal, regular sequence of $\mathfrak{A}$.

We will let $\mathrm{gr}(\mathfrak{A})$ denote the grade of $\mathfrak{A}$. Clearly, $\mathrm{gr}(\mathfrak{A}) = 0$ if and only if $\mathfrak{A} \subseteq Z(R)$. In particular, if R is an integral domain and $\mathfrak{A} \neq (0)$, then $\mathrm{gr}(\mathfrak{A}) \geq 1$. For example, if $R = \mathbb{Z}$, the integers, and $\mathfrak{A} = (x)$, $x \geq 1$, then $\mathrm{gr}(\mathfrak{A}) = 1$. If $\mathfrak{M} = (X_1, \ldots, X_n) \subseteq F[X_1, \ldots, X_n]$, then $\mathrm{gr}(\mathfrak{M}) = n$. We will be interested primarily in ideals of grade at least two. If $\mathrm{gr}(\mathfrak{A}) \geq 2$, then there exists a regular sequence ξ of $\mathfrak{A}$ whose length is at least two.

We have seen in the text that any R-module homomorphism $f: R^{n-1} \mapsto R^n$ can be represented by an $n \times (n-1)$ matrix A [in $M_{n\times(n-1)}(R)$] in the following way: Let $\{\varepsilon_1, \ldots, \varepsilon_{n-1}\}$ and $\{\varepsilon_1', \ldots, \varepsilon_n'\}$ be the canonical bases of R^{n-1} and R^n, respectively. If $f(\varepsilon_i) = \sum_{j=1}^{n} a_{ji}\varepsilon_j'$ for $i = 1, \ldots, n-1$, then set $A = (a_{ji}) \in M_{n\times(n-1)}(R)$. The homomorphism f is then given by matrix multiplication: $f(\lambda) = A\lambda$ for all $\lambda \in R^{n-1}$. We have seen in Theorem 5.36b that f is injective if and only if $\mathrm{Ann}_R(I_{n-1}(A)) = (0)$. We will need the following application of this result.

Lemma D.3 Let $f: R^{n-1} \mapsto R^n$ be an injective, R-module homomorphism. Suppose t is a regular element of R, and set $Q = R^n/f(R^{n-1})$. Let $\mu_t: Q \mapsto Q$ denote the R-module homomorphism given by multiplication by t. Then the following statements are equivalent:

(a) $\mu_t: Q \mapsto Q$ is injective.

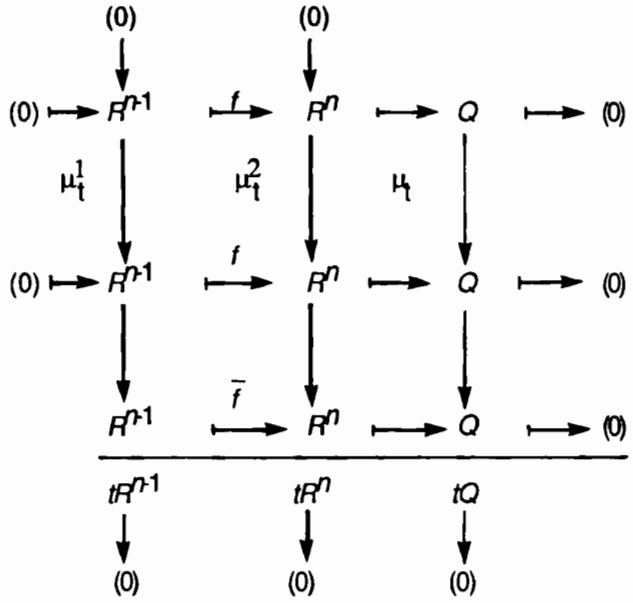
(b) $\bar{f}: R^{n-1}/tR^{n-1} \mapsto R^n/tR^n$ is injective.

(c) $Rt = Rt : I_{n-1}(A)$.

Proof. Before giving a proof of this lemma, let us discuss the notation being used here. Since $f: R^{n-1} \mapsto R^n$ is an R-module homomorphism, f induces an R-module homomorphism $\bar{f}: R^{n-1}/tR^{n-1} \mapsto R^n/tR^n$ given by $\bar{f}(\lambda + tR^{n-1}) = f(\lambda) + tR^n$ for all $\lambda \in R^{n-1}$. In (b), $\bar{f}$ denotes this induced map. In (c), A is the $n \times (n-1)$ matrix representation of f discussed in the previous paragraph.

Since Q is the cokernel of the map f, we have the following commutative diagram:

D.4



The maps μ_t^1 μ_t^2, and μ_t in diagram D.4 are all R-module homomorphisms given by multiplication by t. The unmarked arrows in D.4 are all natural homomorphisms onto the appropriate quotients. In particular, the five-term sequences making up the first two rows and first two columns of D.4 are short exact sequences. By the Snake Lemma [B ; Proposition 2, p. 6],

D.5 $$\mathrm{Ker}(\mu_t^1) \mapsto \mathrm{Ker}(\mu_t^2) \mapsto \mathrm{Ker}(\mu_t) \mapsto \mathrm{Coker}(\mu_t^1) \xrightarrow{\bar{f}} \mathrm{Coker}(\mu_t^2) \mapsto \mathrm{Coker}(\mu_t)$$

is an exact sequence of R-modules. Since t is a regular element of R, μ_t^1 and μ_t^2 are injective R-module homomorphisms. Thus, the exact sequence in D.5 implies $\mathrm{Ker}(\mu_t) \cong \mathrm{Ker}(\bar{f})$. This last isomorphism shows (a) and (b) in the lemma are equivalent.

For any $r \geq 1$, $R^r/tR^r \cong (R/tR)^r$. Let $\bar{A} = (\bar{a}_{ji}) \in M_{n\times(n-1)}(R/tR)$ denote the image of A. Thus, if $A = (a_{ji})$, then $\bar{A} = (\bar{a}_{ji})$ where $\bar{a}_{ji} = a_{ji} + tR \in$

R/tR for all j and i. Then for every $\bar{\lambda} \in R^{n-1}/tR^{n-1}$, $\bar{f}(\bar{\lambda}) = \bar{A}\bar{\lambda}$. Theorem 5.36b implies $\bar{f}$ is injective if and only if $\text{Ann}_{R/tR}(I_{n-1}(\bar{A})) = (0)$. This last equation is in turn equivalent to $Rt : I_{n-1}(A) = Rt$. Thus, (b) and (c) are equivalent. ∎

Let $\mathfrak{A}$ denote a proper ideal in R. Suppose $\mathfrak{A}$ admits a short exact sequence of R-modules of the following form:

D.6 $$(0) \mapsto R^{n-1} \overset{f}{\mapsto} R^n \overset{g}{\mapsto} \mathfrak{A} \mapsto (0)$$

We have seen in the discussion above that there exists an $n \times (n-1)$ matrix $A = (a_{ji}) \in M_{n\times(n-1)}(R)$ such that $f(\lambda) = A\lambda$ for all $\lambda \in R^{n-1}$. The entries in A are defined by the following equations: $f(\varepsilon_i) = \sum_{j=1}^{n} a_{ji}\varepsilon_j'$ for all $i = 1, \ldots, n-1$. If $g(\varepsilon_j') = a_j$ for $j = 1, \ldots, n$, then the $1 \times n$ matrix $B = (a_1, \ldots, a_n) \in M_{1\times n}(R)$ is a matrix representation of g. Thus $g(\xi) = B\xi$ for all $\xi \in R^n$. When dealing with a short exact sequence like the one in D.6, we will replace f and g with A and B, respectively. Thus, D.6 will be written in the following way:

D.7 $$(0) \mapsto R^{n-1} \overset{A}{\mapsto} R^n \overset{B}{\mapsto} \mathfrak{A} \mapsto (0)$$

The R-module homomorphisms in D.7 are given by $\lambda \mapsto A\lambda$ and $\xi \mapsto B\xi$. Since the sequence in D.7 is exact, we have $\text{Ker}(A) = \text{NS}(A) = (0)$, $BA = 0$, and $\mathfrak{A} = Ra_1 + \cdots + Ra_n$. In particular, $\mathfrak{A}$ is generated by the entries of B, and $\mathfrak{A} \neq (0)$.

Using this notation, the Hilbert-Burch theorem can be stated as follows:

Theorem D.8 (Hilbert-Burch) Let R be a Noetherian, integral domain, and let $\mathfrak{A}$ be a proper ideal of R. Suppose $\mathfrak{A}$ has the following two properties:

(a) There exists a short exact sequence of R-modules

$$(0) \mapsto R^{n-1} \overset{A}{\mapsto} R^n \overset{B}{\mapsto} \mathfrak{A} \mapsto (0)$$

for some $A \in M_{n\times(n-1)}(R)$ and $B \in M_{1\times n}(R)$.

(b) $\text{gr}(\mathfrak{A}) \geq 2$.

Then $\mathfrak{A} = I_{n-1}(A)$.

Proof. $A = \begin{bmatrix} a_{11}, \ldots, a_{1n-1} \\ \vdots \quad\quad \vdots \\ a_{n1}, \ldots, a_{nn-1} \end{bmatrix}$ and $B = (a_1, \ldots, a_n)$

We have noted above that $\mathfrak{A} = Ra_1 + \cdots + Ra_n$ and $\mathfrak{A} \neq (0)$. In particular, some a_j is a nonzero element of R. Since $BA = O$, the vector B^t is a non-trivial solution to the homogeneous system of equations $A^tX = O$. Let $\Delta(1, \ldots, \hat{i}, \ldots, n; 1, \ldots, n-1)$ denote the $(n-1) \times (n-1)$ minor of A obtained by deleting the ith row of A. Set $d_i = (-1)^i\Delta(1, \ldots, \hat{i}, \ldots, n; 1, \ldots, n-1)$ for $i = 1, \ldots, n$. Then $I_{n-1}(A) = \sum_{i=1}^{n} Rd_i$.

For each $k \in \{1, \ldots, n-1\}$, let B_k denote the following $n \times n$ matrix:

$$\textbf{D.9} \qquad B_k = \begin{bmatrix} a_{11}, \ldots, a_{1k}, a_{1k}, a_{1k+1}, \ldots, a_{1n-1} \\ \cdot \quad\quad \cdot \quad \cdot \quad \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \quad \cdot \quad \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \quad \cdot \quad \cdot \quad\quad \cdot \\ a_{n1}, \ldots, a_{nk}, a_{nk}, a_{nk+1}, \ldots, a_{nn-1} \end{bmatrix}$$
$$k \quad k+1$$

The integers k and $k+1$ below B_k indicate what column has been repeated. Using Laplace's expansion down the kth column of B_k, we have $0 = \det B_k = \sum_{i=1}^{n} a_{ik}(-1)^k d_i$ for all $k = 1, \ldots, n-1$. Therefore,

$$\textbf{D.10} \qquad A^t \begin{bmatrix} d_1 \\ \cdot \\ \cdot \\ \cdot \\ d_n \end{bmatrix} = O$$

Since multiplication by A is injective, $\mathrm{Ann}_R(I_{n-1}(A)) = (0)$. Therefore, $\mathrm{rk}(A) = n-1$. In particular, $\mathrm{rk}(A^t) = n-1$. Let K denote the quotient field of R. Then $\dim_K(NS(A^t)) = 1$. Consequently, the vectors $(a_1, \ldots, a_n)^t$ and $(d_1, \ldots, d_n)^t$ are linearly dependent over K. Put another way, we have

$$\textbf{D.11} \qquad \mathrm{rank}_K \begin{bmatrix} d_1, \ldots, d_n \\ a_1, \ldots, a_n \end{bmatrix} = 1$$

In particular, every 2×2 minor of the matrix in D.11 is zero. Thus,

D.12 $a_id_j = a_jd_i$ for all $i, j = 1, \ldots, n$.

Since $\mathrm{rk}(A) = n - 1$, some d_i is not zero. Suppose $d_j \neq 0$. Then equation D.12 implies $a_i = (a_j d_j^{-1})d_i$ for all $i = 1, \ldots, n$. Since $\mathfrak{A} \neq (0)$, the quotient $(a_j d_j^{-1})$ is not zero. To simplify notation, let us write $(a_j d_j^{-1}) = r/s$ where $r, s \in R^*$. Then $rd_i \in Rs$ for all $i = 1, \ldots, n$. Therefore, $r \in Rs : I_{n-1}(A)$.

Now R is an integral domain, and $s \neq 0$. In particular, multiplication by s induces an injective R-module homomorphism $\mu_s : \mathfrak{A} \mapsto \mathfrak{A}$. Since $\mathfrak{A} = R^n/AR^{n-1}$, Lemma D.3 implies $Rs = Rs : I_{n-1}(A)$. Hence, $r = us$ for some $u \in R^*$. In particular, $a_i = ud_i$ for all $i = 1, \ldots, n$.

Suppose we can show u is a unit in R. Then

$$\mathfrak{A} = \sum_{i=1}^{n} Ra_i = \sum_{i=1}^{n} Rud_i = \sum_{i=1}^{n} Rd_i = I_{n-1}(A)$$

Hence, the theorem is proved. We finish the proof by showing $\mathrm{gr}(\mathfrak{A}) \geq 2$ implies $u \in U(R)$.

Since $a_i = ud_i$ for $i = 1, \ldots, n$, multiplication by u induces a surjective R-module homomorphism $\mu_u : I_{n-1}(A) \mapsto \mathfrak{A}$. Thus, multiplication by $u^{-1}(\in K)$ induces an R-module homomorphism from $\mathfrak{A}$ to $I_{n-1}(A)$. Since $\mathrm{gr}(\mathfrak{A}) \geq 2$, $\mathfrak{A}$ contains elements x and y such that $(x, y)^t$ is a regular sequence of length two. Therefore, x/u and y/u are elements in $I_{n-1}(A) \subseteq R$. Write $x/u = a$ and $y/u = b$ for some $a, b \in R$. Then $x = ua$ and $y = ub$. In particular, $ay = uab = xb$. Since $(x, y)^t$ is a regular sequence of R, y is not a zero-divisor modulo x. Hence, $a = vx$ for some $v \in R$. Then $bx = ay = vxy$. Since $x \neq 0$, $b = vy$. Therefore, $y = ub = uvy$. Since $y \neq 0$, $uv = 1$. Thus $u \in U(R)$, and the proof of Theorem D.8 is complete. ∎

The Hilbert-Burch theorem gives sufficient conditions for an ideal $\mathfrak{A}$ generated by n elements to be generated by the maximal size minors of some $n \times (n - 1)$ matrix A with entries from R. Ideals of the form $I_t(C)$ with $C \in M_{m \times n}(R)$ and $1 \leq t \leq \min\{m,n\}$ are called determinantal ideals. The study of determinantal ideals is a very active area of research in commutative ring theory. The Hilbert-Burch theorem provides an important class of determinantal ideals, i.e., those ideals generated by the maximal size minors of some $n \times (n - 1)$ matrix.

There is a special version of Theorem D.8 which is used in algebraic geometry. In this version, the integral domain R is a polynomial ring $F[X_0, \ldots, X_n]$ in $n + 1$ indeterminates $X_0, \ldots, X_n$ over a field F. Let $R = F[X_0, \ldots, X_n]$, and let $\mathfrak{A}$ denote a proper ideal of R. We will let $\mathrm{ht}(\mathfrak{A})$ denote the height of $\mathfrak{A}$, $\mathrm{pd}_R(\mathfrak{A})$ the projective dimension of $\mathfrak{A}$, and $\nu(\mathfrak{A})$ the minimal number of generators of $\mathfrak{A}$. The Hilbert-Burch theorem in this context says the following:

Corollary D.13 Let $R = F[X_0, \ldots, X_n]$. Suppose $\mathfrak{A}$ is a homogeneous ideal of R which satisfies the following three properties:

(a) $\mathrm{ht}(\mathfrak{A}) = 2$

(b) $\mathrm{pd}_R(\mathfrak{A}) = 1$

(c) $\nu(\mathfrak{A}) = n + 1$

Then $\mathfrak{A}$ is generated by the maximal size minors of some $(n + 1) \times n$ matrix whose entries are homogeneous polynomials in R.

The proof of Corollary D.13 is beyond the level of this text. However, for readers who have had a course in commutative ring theory, Corollary D.13 follows readily from Theorem D.8. We invite the reader to consult [G] for applications of the Hilbert-Burch theorem in algebraic geometry.

REFERENCES

[B] N. Bourbaki, *Elements of Mathematics—Commutative Algebra*, Chapters 1–7, Springer-Verlag, New York, 1989.

[G] A. Geramita and E. Kani, The Hilbert-Burch theorem, The Curves Seminar at Queens, Queen's Papers in Pure and Applied Mathematics, No. 83.

[K] I. Kaplansky, Commutative Rings, Allyn & Bacon, Boston, 1970.

Notation

T	an arbitrary associative ring with unit	1
T^*	the nonzero elements of T	1
$Z(T)$	the zero divisors of T	1
R	a commutative ring	1
$U(T)$	the units of T	1
$x \sim y$	x and y are associates	1
$x \mid y$	x divides y	2
$M_{n\times n}(F)$	$n \times n$ matrices with entries from F	2
$\det(A)$	the determinant of the matrix A	2
E_{ij}	the i,jth matrix unit	2
$\mathbb{Z}$	the integers	2
F	an arbitrary field	2
$\mathbb{Q}$	the rational numbers	2

$F[X_1, \ldots, X_n]$	polynomials in $X_1, \ldots, X_n$ with coefficients in F	2
$\subseteq$	inclusion	2
$\varnothing$	the empty set	3
$X < Y$	X is strictly contained in Y	3
$J(T)$	The Jacobson radical of T	3
$\mathfrak{L}(T)$	the set of proper left ideals of T	3
PIR	principal ideal ring	4
PID	principal ideal domain	4
$\mathbb{Z}[i]$	Gaussian integers	4
$\hat{\mathbb{Z}}_p$	p-adic integers	4
$F[[X]]$	formal power series in X with coefficients from F	4
$\mathbb{Z}/n\mathbb{Z}$	the integers modulo n	5
$M_{m\times n}(R)$	$m \times n$ matrices with entries from R	5
$\dim_F(V)$	the dimension of V	6
$R[X_1, \ldots, X_n]$	polynomials in $X_1, \ldots, X_n$ with coefficients from R	6
R^n	column vectors of size n with entries from R	6
$(x_1, \ldots, x_n)$	a row vector of size n	6
$(x_1, \ldots, x_n)^t$	a column vector of size n	6
$\boldsymbol{\varepsilon} = \{\varepsilon_1, \ldots, \varepsilon_n\}$	the canonical basis of R^n	6
$\mathrm{Ann}_R(m)$	the annihilator of m	7
$\mathrm{Ann}_R(M)$	the annihilator of M	7
$Z(M)$	the zero divisors of M	7
$\mathrm{Hom}_R(M,N)$	the set of all R-module homomorphisms from M to N	7
$M \cong N$	M is isomorphic to N	7

$\mathrm{Ker}(f)$	the kernel of f	7
$\mathrm{Im}(f)$	the image of f	7
$[A]_{ij}$	the i,jth entry of the matrix A	10
O	the zero matrix	10
A^t	the transpose of A	11
$\mathrm{Row}_i(A)$	the ith row of A	11
$\mathrm{Col}_i(A)$	the ith column of A	11
$A = (\lambda_1; \ldots; \lambda_m)$	the row partition of A	12
$A = (\delta_1 \mid \cdots \mid \delta_n)$	the column partition of A	12
$\mathrm{CS}(A)$	the column space of A	13
$\mathrm{RS}(A)$	the row space of A	13
S_n	the set of all permutations on n letters	14
$\mathrm{sgn}(\sigma)$	the sign of the permutation σ	14
$\Delta(i_1, \ldots, i_t; j_1, \ldots, j_t)$	the $t \times t$ minor of a matrix formed from rows $i_1, \ldots, i_t$ and columns $j_1, \ldots, j_t$	14
$M_{ij}(A)$	the i,jth minor of A	15
$\mathrm{cof}_{ij}(A)$	the i,jth cofactor of A	15
δ_{ij}	the Kronecker delta function	15
$\mathrm{adj}(A)$	the adjoint of A	16
I_n	the $n \times n$ identity matrix	16
$\mathrm{Gl}(n,R)$	the set of $n \times n$ invertible matrices in $M_{n\times n}(R)$	16
$\mathrm{Diag}(d_1, \ldots, d_r)$	a diagonal matrix	17
$C(T)$	the center of the ring T	18
$(M_{n\times n}(R))[X]$	polynomials in X, coefficients from $M_{n\times n}(R)$	18
$\mathrm{Tr}(A)$	the trace of the matrix A	18

$\mathrm{Sl}(n,R)$	$n \times n$ matrices in $M_{n\times n}(R)$ having determinant one	19
$\mathfrak{I}(R)$	the set of ideals of R	23
$x \circ y$	$x + y - xy$	24
$\mathbb{H}$	the real quaterions	27
$\mathbb{C}$	the complex numbers	27
$\bar{z}$	the conjugate of z	27
$\mathbb{R}$	the field of real numbers	27
$I_t(A)$	the ideal generated by all $t \times t$ minors of A	28
$\mathrm{rk}(A)$	the rank of the matrix A	30
$\mathrm{rank}_F(A)$	the classical rank of A over a field F	32
$Q(R)$	the total quotient ring of R	40
$\mathrm{rank}(M)$	the rank of M	41
$I_m(A \mid B)^*$	$m \times m$ minors of A using column B	46
μ_A	left multiplication by the matrix A	49
Γ^c	the complement of the set Γ	53
$\sqrt{\mathfrak{A}}$	the radical of $\mathfrak{A}$	54
$\mathrm{V}(\mathfrak{A})$	the set of prime ideals containing $\mathfrak{A}$	54
$S^{-1}R$	the localization of R at S	60
$T[X]$	polynomials in X with coefficients in T	62
$\partial(f)$	the degree of f	63
$\mathbb{N}_0$	$\{0,1,2,3, \ldots\}$	63
$f_R(z)$	the right evaluation of $f(X)$ at z	66
$f_L(z)$	the left evaluation of $f(X)$ at z	66

$C_A(X)$	the characteristic polynomial of A	69
$\mathrm{Com}(f)$	the companion matrix of f	69
ϑ_A	the R-algebra homomorphism from $R[X]$ to $M_{n\times n}(R)$ given by $f(X) \mapsto f(A)$	71
$R[A]$	the image of ϑ_A	71
N_A	the null ideal of A	71
$\mathrm{NS}(A)$	the null space of A	71
$m_A(X)$	the minimal polynomial of A	75
$\sum c_{\alpha(1)\alpha(2)\cdots\alpha(n)} X_1^{\alpha(1)}X_2^{\alpha(2)}\cdots X_n^{\alpha(n)}$	a polynomial in $X_1, \ldots, X_n$	79
$\mathcal{S}(u_0, \ldots, u_n, v_0, \ldots, v_m)$	Sylvester's matrix	83
$S(u_0, \ldots, u_n, v_0, \ldots, v_m)$	Sylvester's determinant	84
$\mathfrak{R}(f,g)$	the resultant of f and g	85
$\omega(M)$	the weight of M	89
$\partial_{X_i}(f)$	the degree of f with respect to X_i	93
$\mathfrak{R}_{X_i}(f,g)$	the resultant of f and g with respect to X_i	93
h_f	the homogenization of f	103
$\mathfrak{A} : \mathfrak{B}$	the quotient of two ideals	109
$\mathrm{R}_{\mathfrak{P}}$	the localization of R at $\mathfrak{P}$	119
$\mathrm{Bil}_R(M \times N, P)$	the set of all R-bilinear mappings from $M\times N$ to P	136
$\langle\ ,\ \rangle$	the inner product	136
$M \otimes_R N$	the tensor product of M and N	140
$m \otimes n$	the tensor product of m and n	140
$M_1 \otimes_R \cdots \otimes_R M_n$	the tensor product of $M_1, \ldots, M_n$	142

$m_1 \otimes \cdots \otimes m_n$	the tensor product of $m_1, \ldots, m_n$	142
$f \otimes g$	the tensor product of the R-module homomorphisms f and g	143
$\mathrm{Mul}_R(M_1 \times \cdots \times M_n, P)$	the set of all R-multilinear mappings from $M_1 \times \cdots \times M_n$ to P	146
$\mathfrak{F}_k(M)$	the kth Fitting ideal of M	157
M^n	the direct sum of M (n times)	158
$S^{-1}M$	the localization of M at S	161
$M(C)$	the module associated with the matrix C	163
$\mathfrak{T}(M)$	the torsion submodule of M	163
$\mathrm{pd}_R(M) \leq 1$	the projective dimension of M is less than or equal to one	167
$\mathrm{a} \equiv \mathrm{b} \bmod \mathfrak{A}$	$a - b \in \mathfrak{A}$	176
$A \approx B$	A is equivalent to B	184
$d_1 \mid d_2 \mid \cdots \mid d_r$	d_i divides d_{i+1} for all $i = 1, \ldots, r-1$	184
$\mu_R(M)$	the minimal number of generators of M	190
$\mathfrak{D}(\mathrm{A})$	a sequence of invariant factors of A	195
$\mathfrak{E}(\mathrm{A})$	a sequence of elementary divisors of A	196
$A \tilde{s} B$	A is similar to B	205
$\mathcal{S}(A)$	the spectrum of A	214
$E(d)$	the eigenspace associated with d	214
$\mathfrak{R}(A)$	the roots of $C_A(X)$ in R	216
$S(d)$	a free basis of $E(d)$	221
$(A, \cdot)$	a partially ordered set	230

$\mathcal{P}(\Gamma)$	set of all subsets of Γ	231
$\rho : T \mapsto \mathrm{Hom}_{\mathbf{Z}}(M,M)$	a representation of T	235
μ_t^L	left multiplication by t	235
μ_t^R	right multiplication by t	235
f^σ	a specialization of f	244
$\mathbb{P}_K^n$	projective n-space over K	253
$(x_1 : \cdots : x_{n+1})$	a point in $\mathbb{P}_K^n$	253
$V(f)$	the zeros of f	254
order(Γ)	the order of Γ	255
L_A	the homogeneous, linear transformation of $\mathbb{P}_K^2$	255
$i(\Gamma,\Delta;P_j)$	the intersection multiplicity of Γ and Δ at P_j	258
coker(f)	the cokernel of f	262
ht($\mathfrak{A}$)	the height of $\mathfrak{A}$	265
$\mathrm{pd}_R(\mathfrak{A})$	the projective dimension of $\mathfrak{A}$	265
$\nu(\mathfrak{A})$	the minimum number of generators of $\mathfrak{A}$	265

References

1. M. F. Atiyah and I. G. Macdonald, *Introduction to Commutative Algebra*, Addison-Wesley, Reading, Massachusetts, 1969.
2. W. C. Brown, *A Second Course in Linear Algebra*, John Wiley & Sons, New York, 1988.
3. W. C. Brown, *Matrices and Vector Spaces*, Marcel Dekker, New York, 1991.
4. P. Camion., L. S. Levy, and H. B. Mann, Linear equations over a commutative ring, J. Algebra, *18*(3) (1971), 432–446.
5. N. Jacobson, *Basic Algebra I*, W. H. Freeman, New York, 1985.
6. N. Jacobson, *Basic Algebra II*, W. H. Freeman, New York, 1985.
7. J. Lipman, On the Jacobian ideal of the module of differentials, Proc. Am. Math. Soc. *21* (1969), 422–426.
8. R. E. MacRae, On the application of Fitting invariants, J. Algebra *2*(2) (1965), 153–169.
9. R. Bruce Richter and William P. Wardlaw, Diagonalization over commutative rings, Am. Math. Monthly *97*(3) (1990), 223–227.
10. O. Zariski and P. Samuel, *Commutative Algebra*, Vol. I, Van Nostrand, New York, 1958 (new edition Springer, 1975).

BIBLIOGRAPHY

I. Kaplansky, Elementary divisors and modules, Trans. Am. Math. Soc., *66*(1949), 464–491.

M. Marcus, *Introduction to Modern Algebra,* Marcel Dekker, New York, 1978.

N. McCoy, *Rings and Ideals*, Carus Math. Monogr. *8*, Mathematical Association of America.

Index