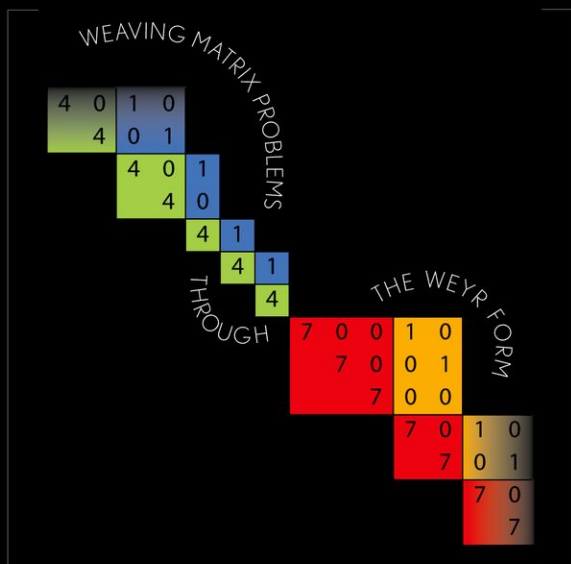


ADVANCED TOPICS IN LINEAR ALGEBRA



Kevin C. O'Meara ■ John Clark ■ Charles I. Vinsonhaler

Advanced Topics in Linear Algebra

This page intentionally left blank

Advanced Topics in Linear Algebra

Weaving Matrix Problems through the Weyr Form

KEVIN C. O'MEARA

JOHN CLARK

CHARLES I. VINSONHALER

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2011 by Oxford University Press

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016
www.oup.com

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data

O'Meara, Kevin C.

Advanced topics in linear algebra : weaving matrix problems through the
Weyr Form / Kevin C. O'Meara, John Clark, Charles I. Vinsonhaler.
p. cm. Includes bibliographical references and index.

ISBN 978-0-19-979373-0

1. Algebras, Linear. I. Clark, John. II. Vinsonhaler, Charles Irvin, 1942- III. Title.

QA184.2.O44 2011

512'.5-dc22 2011003565

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

DEDICATED TO

Sascha, Daniel, and Nathania
Kevin O'Meara

Austina and Emily Grace Clark
John Clark

Dorothy Snyder Vinsonhaler
Chuck Vinsonhaler

This page intentionally left blank

CONTENTS

Preface	xi
Our Style	xvii
Acknowledgments	xxi

PART ONE: The Weyr Form and Its Properties 1

1. Background Linear Algebra	3
1.1. The Most Basic Notions	4
1.2. Blocked Matrices	11
1.3. Change of Basis and Similarity	17
1.4. Diagonalization	22
1.5. The Generalized Eigenspace Decomposition	27
1.6. Sylvester's Theorem on the Matrix Equation $AX - XB = C$	33
1.7. Canonical Forms for Matrices	35
Biographical Notes on Jordan and Sylvester	42
2. The Weyr Form	44
2.1. What Is the Weyr Form?	46
2.2. Every Square Matrix Is Similar to a Unique Weyr Matrix	56
2.3. Simultaneous Triangularization	65
2.4. The Duality between the Jordan and Weyr Forms	74
2.5. Computing the Weyr Form	82
Biographical Note on Weyr	94
3. Centralizers	96
3.1. The Centralizer of a Jordan Matrix	97
3.2. The Centralizer of a Weyr Matrix	100
3.3. A Matrix Structure Insight into a Number-Theoretic Identity	105
3.4. Leading Edge Subspaces of a Subalgebra	108

3.5. Computing the Dimension of a Commutative Subalgebra	114
Biographical Note on Frobenius	123
4. The Module Setting	124
4.1. A Modicum of Modules	126
4.2. Direct Sum Decompositions	135
4.3. Free and Projective Modules	144
4.4. Von Neumann Regularity	152
4.5. Computing Quasi-Inverses	159
4.6. The Jordan Form Derived Module-Theoretically	169
4.7. The Weyr Form of a Nilpotent Endomorphism: Philosophy	174
4.8. The Weyr Form of a Nilpotent Endomorphism: Existence	178
4.9. A Smaller Universe for the Jordan Form?	185
4.10. Nilpotent Elements with Regular Powers	188
4.11. A Regular Nilpotent Element with a Bad Power	195
Biographical Note on Von Neumann	197
PART TWO: Applications of the Weyr Form	199
5. Gerstenhaber's Theorem	201
5.1. k -Generated Subalgebras and Nilpotent Reduction	203
5.2. The Generalized Cayley–Hamilton Equation	210
5.3. Proof of Gerstenhaber's Theorem	216
5.4. Maximal Commutative Subalgebras	221
5.5. Pullbacks and 3-Generated Commutative Subalgebras	226
Biographical Notes on Cayley and Hamilton	236
6. Approximate Simultaneous Diagonalization	238
6.1. The Phylogenetic Connection	241
6.2. Basic Results on ASD Matrices	249
6.3. The Subalgebra Generated by ASD Matrices	255
6.4. Reduction to the Nilpotent Case	258
6.5. Splittings Induced by Epsilon Perturbations	260
6.6. The Centralizer of ASD Matrices	265
6.7. A Nice 2-Correctable Perturbation	268
6.8. The Motzkin–Taussky Theorem	271
6.9. Commuting Triples Involving a 2-Regular Matrix	276
6.10. The 2-Regular Nonhomogeneous Case	287
6.11. Bounds on $\dim \mathbb{C}[A_1, \dots, A_k]$	297
6.12. ASD for Commuting Triples of Low Order Matrices	301
Biographical Notes on Motzkin and Taussky	307

7. Algebraic Varieties	309
7.1. Affine Varieties and Polynomial Maps	311
7.2. The Zariski Topology on Affine n -Space	320
7.3. The Three Theorems Underpinning Basic Algebraic Geometry	326
7.4. Irreducible Varieties	328
7.5. Equivalence of ASD for Matrices and Irreducibility of $\mathcal{C}(k, n)$	339
7.6. Gerstenhaber Revisited	342
7.7. Co-Ordinate Rings of Varieties	347
7.8. Dimension of a Variety	353
7.9. Guralnick's Theorem for $\mathcal{C}(3, n)$	364
7.10. Commuting Triples of Nilpotent Matrices	370
7.11. Proof of the Denseness Theorem	378
Biographical Notes on Hilbert and Noether	381
 Bibliography	 384
Index	390

This page intentionally left blank

PREFACE

“Old habits die hard.” This maxim may help explain why the Weyr form has been almost completely overshadowed by its cousin, the Jordan canonical form. Most have never even heard of the Weyr form, a matrix canonical form discovered by the Czech mathematician Eduard Weyr in 1885. In the 2007 edition of the *Handbook of Linear Algebra*, a 1,400-page, authoritative reference on linear algebra matters, there is *not a single mention* of the Weyr form (or its associated Weyr characteristic). But this canonical form is as useful as the Jordan form, which was discovered by the French mathematician Camille Jordan in 1870. Our book is in part an attempt to remedy this unfortunate situation of a grossly underutilized mathematical tool, by making the Weyr form more accessible to those who use linear algebra at its higher level. Of course, that class includes most mathematicians, and many others as well in the sciences, biosciences, and engineering. And we hope our book also helps popularize the Weyr form by demonstrating its topical relevance, to both “pure” and “applied” mathematics. We believe the applications to be interesting and surprising.

Although the unifying theme of our book is the development and applications of the Weyr form, this does not adequately describe the full scope of the book. The three principal applications—to matrix commutativity problems, approximate simultaneous diagonalization, and algebraic geometry—bring the reader right up to current research (as of 2010) with a number of open questions, and also use techniques and results in linear algebra not involving canonical forms. And even in topics that are familiar, we present some unfamiliar results, such as improving on the known fact that commuting matrices over an algebraically closed field can be simultaneously triangularized.

Matrix canonical forms (with respect to similarity) provide exemplars for each similarity class of square $n \times n$ matrices over a fixed field. Their aesthetic qualities have long been admired. But canonical forms also have some very concrete applications. The authors were drawn to the Weyr form through a

question that arose in phylogenetic invariants in biomathematics in 2003. Prior to that, we too were completely unaware of the Weyr form. It has been a lot of fun rediscovering the lovely properties of the Weyr form and, in some instances, finding new properties. In fact, quite a number of results in our book have (apparently) not appeared in the literature before. There is a wonderful mix of ideas involved in the description, derivation, and applications of the Weyr form: linear algebra, of course, but also commutative and noncommutative ring theory, module theory, field theory, topology (Euclidean and Zariski), and algebraic geometry. We have attempted to blend these ideas together throughout our narrative. As much as possible, given the limits of space, we have given self-contained accounts of the nontrivial results we use from outside the area of linear algebra, thereby making our book accessible to a good graduate student. For instance, we develop from scratch a fair bit of basic algebraic geometry, which is unusual in a linear algebra book. If nothing else, we claim to have written quite a novel linear algebra text. We are not aware of any book with a significant overlap with the topics in ours, or of any book that devotes an entire chapter to the Weyr form. However, Roger Horn recently informed us (in September 2009) that the upcoming second edition of the Horn and Johnson text *Matrix Analysis* will have a section on the Weyr form in Chapter 3. Of course, whether our choice of topics is good or bad, and what sort of job we have done, must ultimately be decided by the reader.

All seven chapters of our book begin with a generous introduction, as do most sections within a chapter. We feel, therefore, that there is not a lot of point in describing the chapter contents within this preface, beyond the barest summary that follows. Besides, the reader is not expected to know what the Weyr form is at this time.

PART I: THE WEYR FORM AND ITS PROPERTIES

1: **Background Linear Algebra**

We do a quick run-through of some of the more important basic concepts we require from linear algebra, including diagonalization of matrices, the description of the Jordan form, and desirable features of canonical matrix forms in general.

2: **The Weyr Form**

Here we derive the Weyr form from scratch, establish its basic properties, and detail an algorithm for computing the Weyr form of nilpotent matrices (always the core case). We also derive an important duality between the Jordan and Weyr structures of nilpotent matrices.

3: **Centralizers**

The matrices that centralize (commute with) a given nilpotent Jordan matrix have a known explicit description. Here we do likewise for the

Weyr form, for which the centralizer description is simpler. It is this property that gives the Weyr form an edge over its Jordan counterpart in a number of applications.

4: **The Module Setting**

The Jordan form has a known ring-theoretic derivation through decompositions of finitely generated modules over principal ideal domains. In this chapter we derive the Weyr form ring theoretically, but in an entirely different way, by using ideas from decompositions of projective modules over von Neumann regular rings. The results suggest that the Weyr form lives in a somewhat bigger universe than its Jordan counterpart, and is perhaps more natural.

PART II: APPLICATIONS OF THE WEYR FORM

5: **Gerstenhaber's Theorem**

The theorem states that the subalgebra $F[A, B]$ generated by two commuting $n \times n$ matrices A and B over a field F has dimension at most n . It was first proved using algebraic geometry, but later Barriá and Halmos, and Laffey and Lazarus, gave proofs using only linear algebra and the Jordan form. Here we simplify the Barriá–Halmos proof even further through the use of the Weyr form in tandem with the Jordan form, utilizing an earlier duality.

6: **Approximate Simultaneous Diagonalization**

Complex $n \times n$ matrices A_1, A_2, \dots, A_k are called approximately simultaneously diagonalizable (ASD) if they can be perturbed to simultaneously diagonalizable matrices B_1, B_2, \dots, B_k . In this chapter we attempt to show how the Weyr form is a promising tool (more so than the Jordan form) for establishing ASD of various classes of commuting matrices using explicit perturbations. The ASD property has been used in the study of phylogenetic invariants in biomathematics, and multivariate interpolation.

7: **Algebraic Varieties**

Here we give a largely self-contained account of the algebraic geometry connection to the linear algebra problems studied in Chapters 5 and 6. In particular, we cover most of the known results on the irreducibility of the variety $\mathcal{C}(k, n)$ of all k -tuples of commuting complex $n \times n$ matrices. The Weyr form is used to simplify some earlier arguments. Irreducibility of $\mathcal{C}(k, n)$ is surprisingly equivalent to all k commuting complex $n \times n$ matrices having the ASD property described in Chapter 6. But a number of ASD results are known only through algebraic geometry. Some of this work is quite recent (2010).

Our choice of the title *Advanced Topics in Linear Algebra* indicates that we are assuming our reader has a solid background in undergraduate linear algebra (see the introduction to Chapter 1 for details on this). However, it is probably fair to say that our treatment is at the higher end of “advanced” but without being comprehensive, compared say with Roman’s excellent text *Advanced Linear Algebra*,¹ in the number of topics covered. For instance, while some books on advanced linear algebra might take the development of the Jordan form as one of their goals, we assume our readers have already encountered the Jordan form (although we remind readers of its properties in Chapter 1). On the other hand, we do not assume our reader is a specialist in linear algebra. The book is designed to be read in its entirety if one wishes (there is a continuous thread), but equally, after a reader has assimilated Chapters 2 and 3, each of the four chapters that follow Chapter 3 can be read in isolation, depending on one’s “pure” or “applied” leanings.

At the end of each chapter, we give brief biographical sketches of one or two of the principal architects of our subject. It is easy to forget that mathematics has been, and continues to be, developed by real people, each generation building on the work of the previous—not tearing it down to start again, as happens in many other disciplines. These sketches have been compiled from various sources, but in particular from the MacTutor History of Mathematics web site of the University of St. Andrews, Scotland [<http://www-history.mcs.st-andrews.ac.uk/Biographies>], and I. Kleiner’s *A History of Abstract Algebra*. Note, however, that we have given biographies only for mathematicians who are no longer alive.

When we set out to write this book, we were not thinking of it as a text for a course, but rather as a reference source for teachers and researchers. But the more we got into the project, the more apparent it became that parts of the book would be suitable for graduate mathematics courses (or fourth-year honors undergraduate courses, in the case of the better antipodean universities). True, we have not included exercises (apart from a handful of test questions), but the nature of the material is such that an instructor would find it rather easy (and even fun) to make up a wide range of exercises to suit a tailored course. As to the types of course, a number spring to mind:

- (1) A second-semester course following on from a first-year graduate course in linear algebra, covering parts of Chapters 1, 2, 3, and 6.

1. Apart from our background in Chapter 1, there is no overlap in the topics covered in our book and that of Roman.

- (2) A second-semester course following on from an abstract algebra course that covered commutative and noncommutative rings, covering parts of Chapters 1, 2, 3, 4, 5, and 7.
- (3) The use of Chapter 4 as a supplement in a course on module theory.
- (4) The use of Chapter 7 as a supplement in a course on algebraic geometry or biomathematics (e.g., phylogenetics).

The authors welcome comments and queries from readers. Please use the following e-mail addresses:

<code>staf198@ext.canterbury.ac.nz</code>	(Kevin O'Meara)
<code>jclark@maths.otago.ac.nz</code>	(John Clark)
<code>charles.vinsonhaler@uconn.edu</code>	(Chuck Vinsonhaler)

This page intentionally left blank

OUR STYLE

Mathematicians are expected to be very formal in their writings. An unintended consequence of this is that mathematics has more than its share of rather boring, pedantic, and encyclopedic books—good reading for insomniacs. We have made a conscious decision to write in a somewhat lighter and more informal style. We comment here on some aspects of this style, so that readers will know what to expect. The mathematical content of our arguments, on the other hand, is always serious.

Some mathematics writers believe that because they have formally spelled out all the precise definitions of every concept, often lumped together at the very beginning of a chapter or section, the reader must be able to understand and appreciate their arguments. This is not our experience. (One first expects to see the menu at a restaurant, not a display of all the raw ingredients.) And surely, if a result is stated in its most general form, won't the reader get an even bigger insight into the wonders of the concepts? Mistaken again, in our view, because this may obscure the essence of the result. To complete the trifecta of poor writing, mathematicians sometimes try to tell the reader *everything* they know about a particular topic; in so doing, they often cloud perspective. We have kept the formal (displayed) statements of definitions to a minimum—reserved for the most important concepts. We have also attempted to delay the formal definition until after suitable motivation of the concept. The concept is usually then illustrated by numerous examples. And in the development proper, we don't tell everything we know. In fact, we often invite (even challenge) the reader to continue the exploration, sometimes in a footnote.

We make no apology for the use of whimsy.¹ In our view, there is a place for whimsy even within the erudite writings of mathematicians. It can help put a human face on the authors (we are not high priests) and can energize a reader to continue during the steeper climbs. Our whimsical comments are mostly reserved for an occasional footnote. But footnotes, being footnotes, can be skipped without loss of continuity to the story.

We have tried to avoid the formality of article writing in referencing works. Thus, rather than say “see Proposition 4.8 (2) and the Corollary on p. 222 of [BAII]” we would tend to say simply “see Chapter 4 of Jacobson’s *Basic Algebra II*.” Likewise, an individual paper by Joe Blog that is listed in our bibliography will usually be referred to as “the 2003 paper by Blog,” if there is only one such paper. The interested reader can then consult the source for more detail.

What constitutes “correct grammar” has been a source of much discussion and ribbing among the three authors, prompted by their different education backgrounds (New Zealand, Scotland, and U.S.A.). By and large, the British Commonwealth has won out in the debate. But we are conscious, for example, of the difference in the British and American use of “that versus which,”² and in punctuation. So please bear with us.

Our notation and terminology are fairly standard. In particular, we don’t put brackets around the argument in the dimension $\dim V$ of a vector space V or the rank of a matrix A , $\text{rank } A$. However, we do use brackets if both the mathematical operator and argument are in the same case. Thus, we write $\ker(b)$ and $\text{im}(p)$ for the kernel and image of module homomorphisms b and p , rather than the visually off-putting $\ker b$ and $\text{im } p$. Undoubtedly, there will be some inconsistencies in our implementation of this policy. An index entry such as Joe Blog’s theorem, 247, **256**, 281 indicates that the principal statement or discussion of the theorem can be found on page 256, the one in boldface. This is done sparingly, reserved for the most important concepts, definitions, or results. Very occasionally, an entry may have more than one boldfaced page to indicate the most important, but separate, treatments of a topic.

Finally, a word to a reader who perceives some “cheerleading” on our part when discussing the Weyr form. We have attempted to be even-handed in our

1. In a 2009 interview (by Geraldine Doogue, Australian ABC radio), Michael Palin (of Monty Python and travel documentary fame, and widely acclaimed as a master of whimsy) was asked why the British use whimsy much more so than Americans. His reply, in part, was that Britain has had a more settled recent history. America has been more troubled by wars and civil rights. Against this backdrop, Americans have tended to take things more seriously than the British.

2. Our rule is “that” introduces a defining clause, whereas “which” introduces a nondefining clause.

treatment of the Weyr and Jordan forms (the reader should find ample evidence of this). But when we are very enthused about a particular result or concept, we tell our readers so. Wouldn't life be dull without such displays of feeling? Unfortunately, mathematics writers often put a premium on presenting material in a deadpan, minimalist fashion.

This page intentionally left blank

ACKNOWLEDGMENTS

These fall into two groups : (1) A general acknowledgment of those people who contributed to the mathematics of our book or its publishing, and (2) A personal acknowledgment from each of the three authors of those who have given moral and financial support during the writing of the book, as well as a recognition of those who helped support and shape them as professional mathematicians over some collective 110 years!

We are most grateful to Mike Steel (University of Canterbury, New Zealand) for getting us interested in the linear algebra side of phylogenetics, and to Elizabeth Allman (University of Alaska, Fairbanks) for contributing the section on phylogenetics in Chapter 6. Many thanks to Paul Smith (University of Washington, Seattle) for supplying the proof of a Denseness Theorem in Chapter 7. We also single out Roger Horn (University of Utah, Salt Lake City) for special thanks. His many forthright, informative comments on an earlier draft of our book, and his subsequent e-mails, have greatly improved the final product.

It is a pleasure to acknowledge the many helpful comments, reference sources, technical advice, and the like from other folk, particularly Pere Ara, John Burke, Austina Clark, Herb Clemens, Keith Conrad, Ken Goodearl, Robert Guralnick, John Hannah, John Holbrook, Robert Kruse, James Milne, Ross Moore, Miki Neumann, Keith Nicholson, Vadim Olshevsky, Matja Omladič, Bob Plemmons, John Shanks, Boris Shekhtman, Klemen Šivic, Molly Thomson, Daniel Velleman (Editor of *American Mathematical Monthly*), Graham Wood, and Milen Yakimov.

To the four reviewers who reported to Oxford University Press on an earlier draft of our book, and who made considered, insightful comments, we say thank you very much. In particular, we thank the two of you who suggested that our original title *The Weyr Form: A Useful Alternative to the Jordan Canonical Form* did not convey the full scope of our book.

Finally, our sincere thanks to editor Phyllis Cohen and her assistant Hallie Stebbins, production editor Jennifer Kowing, project manager Viswanath Prasanna from Glyph International, and the rest of the Oxford University Press (New York) production team (especially the copyeditor and typesetter) for their splendid work and helpful suggestions. They freed us up to concentrate on the writing, unburdened by technical issues. All queries from us three greenhorns were happily and promptly answered. It has been a pleasure working with you.

From Kevin O'Meara. The biggest thanks goes to my family, of whom I am so proud: wife Leelalai, daughters Sascha and Nathania, and son Daniel. They happily adopted a new member into the family, “the book.” Thanks also to those who fed and sheltered me during frequent trips across the Tasman (from Brisbane to Christchurch and Dunedin), and across the Pacific (from Christchurch to Storrs, Connecticut), in connection with the book (or its foundations): John and Anna-Maree Burke, Brian and Lynette O'Meara, Lloyd and Patricia Ashby, Chuck and Patty Vinsonhaler, Mike and Susan Stuart, John and Austina Clark, Gabrielle and Murray Gormack. I have had great support from the University of Connecticut (U.S.A.) during my many visits over the last 30 years, particularly from Chuck Vinsonhaler and Miki Neumann. The University of Otago, New Zealand (host John Clark) has generously supported me during the writing of this book. Many fine mathematicians have influenced me over the years: Pere Ara, Richard Brauer, Ken Goodearl, Israel Herstein, Nathan Jacobson, Robert Kruse (my Ph.D. adviser), and James Milne, to name a few. I have also received generous support from many mathematics secretaries and technical staff, particularly in the days before I got round to learning \LaTeX : Ann Tindal, Tammy Prentice, Molly Thomson, and John Spain are just four representative examples. Finally, I thank Gus Oliver for his unstinting service in restoring the health of my computer after its bouts of swine 'flu.

From John Clark. First and foremost, I'm most grateful to Kevin and Chuck for inviting me on board the good ship *Weyr form*. Thanks also to the O'Meara family for their hospitality in Brisbane and to the Department of Mathematics and Statistics of the University of Otago for their financial support. Last, but certainly not least, my love and gratitude to my wife Austina for seeing me through another book.

From Chuck Vinsonhaler. I am grateful to my wife Patricia for her support, and thankful for the mathematical and expository talents of my coauthors.

PART ONE

The Weyr Form and Its Properties

In the four chapters that compose the first half of our book, we develop the Weyr form and its properties, starting from scratch. Chapters 2 and 3 form the core of this work. Chapter 1 can be skipped by readers with a solid background in linear algebra, while Chapter 4, which gives a ring-theoretic derivation of the Weyr form, is optional (but recommended) reading. Applications involving the Weyr form come in Part II.

This page intentionally left blank

Background Linear Algebra

Most books in mathematics have a background starting point. Ours is that the reader has had a solid undergraduate (or graduate) education in linear algebra. In particular, the reader should feel comfortable with an abstract vector space over a general field, bases, dimension, matrices, determinants, linear transformations, change of basis results, similarity, eigenvalues, eigenvectors, characteristic polynomial, the Cayley–Hamilton theorem, direct sums, diagonalization, and has at least heard of the Jordan canonical form. What is important is not so much a *knowledge* of results in linear algebra as an *understanding* of the fundamental concepts. In actual fact, there are rather few specific results needed as a prerequisite for understanding this book.

The way linear algebra is taught has changed greatly over the last 50 years, and mostly for the better. Also, what was once taught to undergraduates is now often taught to graduate students. Prior to around the 1990s, linear algebra was at times presented as the poor cousin to calculus (or analysis). In some circles that view persists, but now most agree that linear algebra rivals calculus in applicability. (Every time one “Googles,” there is a calculation of a principal eigenvector of a gigantic matrix, of order several billion, to determine

page rank. Amazing!¹) This book was also motivated by an application, to phylogenetics, as discussed in Chapter 6.

The authors are of the school that believes not teaching linear transformations in linear algebra courses is to tell only half the story, even if one's primary applications are to matrices. The full power of a linear algebra argument often comes from flipping back and forth between a matrix view and a transformation view. Without linear transformations, many similarity results for matrices lose their full impact. Also, the concept of an invariant subspace of a linear transformation is one of the most central in all linear algebra. (As a special case, the notion of an eigenvector corresponds to a 1-dimensional invariant subspace.)

In this chapter, we will quickly run through a few of the more important basic concepts we require, but not in any great depth, with very few proofs, and sometimes scant motivation. The concepts are covered in many, many texts. The reader who wants more detail may wish to consult his or her own favorites. Ours include the books by Kenneth Hoffman and Ray Kunze (*Linear Algebra*), Roger Horn and Charles Johnson (*Matrix Analysis*), and Keith Nicholson (*Linear Algebra with Applications*). But there are many other fine books on linear algebra. Our advice to a reader who is already comfortable with the basics of linear algebra (as outlined in the opening paragraph) is to proceed directly to Chapter 2, and return only to check on notation, etc., should the need arise.

1.1 THE MOST BASIC NOTIONS

It's time to get down to the nitty-gritty, beginning with a summary of basic notions in linear algebra. In the first few pages, it is hard to avoid the unexciting format of recalling definitions, registering notation, and blandly stating results. In short, the things mathematics books are renowned for? Bear with us—our treatment will lighten up later in the chapter, when we not only recall concepts but also (hopefully) convey our particular slant on them. In basic calculus and analysis, there is probably not a great variation in how two (competent) individuals view the material. The mental pictures are pretty much the same. But it is less clear what is going on inside a linear algebraist's head. The authors would venture to say that there is more variation in how individuals view the subject matter of linear algebra. (For instance, some get by in linear algebra without using linear transformations, although it would be extremely rare for someone in calculus to never use functions.) It often depends on an individual's particular background in other mathematics.

1. See, for example, the 2006 article "The \$25,000,000,000 eigenvector: The linear algebra behind Google," by K. Bryan and T. Liese.

That is not to say one view of linear algebra is right and another is flatly wrong. If there are particular points of view that come across in the present book, their origins probably lie in the time the authors have worked with algebraic structures (such as semigroups, groups, rings, and associative algebras), and with having developed a healthy respect for such disciplines as category theory, universal algebra, and algebraic geometry. Of course, we have also come to admire the beautiful concepts in analysis and topology, some of which we use in Chapters 6 and 7. In this respect, our philosophical view is mainstream—mathematics, of all disciplines, should never be compartmentalized.

The letter F will denote a field, usually **algebraically closed** (such as the complex field \mathbb{C}), that is, every polynomial over F of positive degree has roots in F . The space of all n -tuples (which we usually write as column vectors) of elements from F is denoted by F^n . This space is the model for all n -dimensional vector spaces over F , because every n -dimensional space is isomorphic to F^n . By the **standard basis** for F^n we mean the basis

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}.$$

The ring of polynomials, in the indeterminate x and with coefficients from F , is denoted by $F[x]$. This ring plays a similar role in linear algebra to that of the ring \mathbb{Z} of integers in group theory. (Both rings are Euclidean domains and, for instance, the order of a group element, as an element of the ring \mathbb{Z} , translates to the minimal polynomial of a matrix, as an element of the ring $F[x]$.) If V is a vector space over F (usually finite-dimensional), its dimension is denoted by $\dim V$. The subspace of V spanned (or generated) by vectors v_1, v_2, \dots, v_n is denoted by $\langle v_1, v_2, \dots, v_n \rangle$. If U_1, U_2, \dots, U_k are subspaces of V , their **sum** is the subspace

$$U_1 + U_2 + \dots + U_k = \{v \in V : v = u_1 + u_2 + \dots + u_k \text{ for some } u_i \in U_i\}.$$

For a linear transformation $T : V \rightarrow W$ from one vector space V to another W , the **rank** of T is the dimension of the image $T(V)$, and the **nullity** of T is the dimension of the **null space** or **kernel**, $\ker T = \{v \in V : T(v) = 0\}$. We have the fundamental rank, nullity connection:

$$\text{rank } T + \text{nullity } T = \dim V.$$

For an $m \times n$ matrix A , this translates as

$$\text{rank } A + \text{nullity } A = n = \text{the number of columns of } A,$$

where $\text{nullity } A$ is the dimension of the solution space of the homogeneous system $A\mathbf{x} = \mathbf{0}$, and $\text{rank } A$ is either column rank or row rank (maximum number of independent columns or rows; they are the same). The matrix A is said to have **full column-rank** if $\text{rank } A = n$ (its columns are linearly independent).

The set of all $n \times n$ (square) matrices over F is denoted by $M_n(F)$. The arithmetic of $M_n(F)$ under addition, multiplication, and scalar multiplication is the most natural model of noncommutative (but associative) arithmetic.² This is one of the principal reasons why linear algebra is such a powerful tool.³

So much of linear algebra and its applications revolve around the concepts of eigenvalues and eigenvectors. Our book is no exception. An **eigenvalue** of a matrix $A \in M_n(F)$, or a linear transformation $T : V \rightarrow V$, is a scalar $\lambda \in F$ such that $Av = \lambda v$ or $T(v) = \lambda v$ for some nonzero vector v (in F^n or V , respectively). Any such v is called an **eigenvector** of A or T corresponding to the eigenvalue λ . The **eigenspace** of A corresponding to λ is $E(\lambda) = \ker(\lambda I - A)$, which is just the set of all eigenvectors corresponding to λ together with the zero vector. (Here I is the identity matrix in $M_n(F)$.) By the **geometric multiplicity** of λ we mean the dimension of $E(\lambda)$. The **characteristic polynomial** of A is $p(x) = \det(xI - A)$. Although this polynomial far from “characterizes” the matrix A , it does reflect many of its important properties. For instance, the zeros of $p(x)$ are precisely the eigenvalues of A . The reason why we often restrict F to being algebraically closed is to ensure eigenvalues always exist. In this case, if $\lambda_1, \lambda_2, \dots, \lambda_k$ are the distinct eigenvalues of A and $p(x) = (x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}$ is the factorization of the characteristic polynomial into linear factors, then m_i is called the **algebraic multiplicity** of the eigenvalue λ_i (and $m_1 + m_2 + \cdots + m_k = n$). The geometric multiplicity of an eigenvalue can never exceed its algebraic multiplicity. A frequently used observation is that the eigenvalues of a triangular matrix are its diagonal entries.

2. The so-called Wedderburn–Artin theorem of ring theory more or less confirms this if F is algebraically closed: the $M_n(F)$ are the only simple, finite-dimensional associative algebras, and finite direct products of these algebras give all “well-behaved” finite-dimensional algebras.

3. Another reason for the success of linear algebra, of course, is that it is suited to studying linear approximation problems.

For a polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m \in F[x]$ and square matrix $A \in M_n(F)$, we can form the matrix polynomial

$$f(A) = a_0I + a_1A + a_2A^2 + \cdots + a_mA^m.$$

This polynomial evaluation map $f \mapsto f(A)$ (for a fixed A), from $F[x]$ to $M_n(F)$, is a simple but useful algebra homomorphism (i.e., a linear mapping that preserves multiplication). The **Cayley–Hamilton theorem** says that over any field F , every square matrix A vanishes at its characteristic polynomial: $p(A) = 0$ where p is the characteristic polynomial of A .

The square matrices A that have 0 as an eigenvalue are the singular (noninvertible) matrices, because they are the matrices for which the homogeneous system $A\mathbf{x} = 0$ has a nonzero solution. A square matrix A is called **nilpotent** if $A^r = 0$ for some $r \in \mathbb{N}$ and in this case the least such r is the (**nilpotency**) **index** of A . If A is nilpotent then 0 is its only eigenvalue. (For if $A^r = 0$ and $A\mathbf{x} = \lambda\mathbf{x}$ for some nonzero \mathbf{x} , then $0 = A^r\mathbf{x} = \lambda^r\mathbf{x}$, which implies $\lambda = 0$.) Over an algebraically closed field, the converse also holds, as we show in our first proposition:

Proposition 1.1.1

Over an algebraically closed field, an $n \times n$ matrix A is nilpotent if and only if 0 is the only eigenvalue of A . Also, a square matrix that does not have two distinct eigenvalues must be the sum of a scalar matrix λI and a nilpotent matrix.

Proof

The second statement follows from the first because if λ is the only eigenvalue of A , then 0 is the only eigenvalue of $A - \lambda I$ (and $A = \lambda I + (A - \lambda I)$). Suppose 0 is the only eigenvalue of A . Since the field is algebraically closed, the characteristic polynomial of A must be $p(x) = x^n$. By the Cayley–Hamilton theorem, $0 = p(A) = A^n$ and so A is nilpotent. \square

Note that the argument breaks down (not the Cayley–Hamilton theorem, which holds over any field) if the characteristic polynomial doesn't factor completely. For instance, over the real field \mathbb{R} , the matrix

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

has zero as its only eigenvalue but is not nilpotent. (However $A^3 + A = 0$, consistent with the Cayley–Hamilton theorem.)

Now $M_n(F)$ is not just a vector space under matrix addition and scalar multiplication, but also a ring with identity under matrix addition and

multiplication, with scalar multiplication and matrix multiplication nicely intertwined by the law⁴ $\lambda(AB) = (\lambda A)B = A(\lambda B)$ for all $\lambda \in F$ and all $A, B \in M_n(F)$. In this context, we refer to $M_n(F)$ as an (associative) **algebra** over the field F . (The general definition of an “algebra over a commutative ring” is given in Definition 4.1.8.) By a **subalgebra** of $M_n(F)$ we mean a subset $B \subseteq M_n(F)$ that contains the identity matrix and is closed under scalar multiplication, matrix addition, and matrix multiplication (in other words, a subspace that is also a subring). Given a subset $S \subseteq M_n(F)$, there is a unique smallest subalgebra of $M_n(F)$ containing S , namely, the intersection of all subalgebras containing S . This is called the **subalgebra generated by** S , and is denoted by $F[S]$. In the case where $S = \{A_1, A_2, \dots, A_k\}$ consists of a finite number k of matrices, we say that $F[S]$ is **k -generated (as an algebra)** and write $F[S] = F[A_1, A_2, \dots, A_k]$. For a single matrix $A \in M_n(F)$, clearly

$$F[A] = \{f(A) : f \in F[x]\}.$$

In fact $\{I, A, A^2, \dots, A^{m-1}\}$ is a vector space basis for $F[A]$ if A^m is the first power that is linearly dependent on the earlier powers. Describing the members of $F[A_1, A_2, \dots, A_k]$ when $k > 1$, or even computing the dimension of this subalgebra, is in general an exceedingly difficult problem.

Over an algebraically closed field F , a **nonderogatory** matrix is a square matrix $A \in M_n(F)$, all of whose eigenspaces are 1-dimensional. This is not the same thing as A having n distinct eigenvalues, although the latter would certainly be sufficient. Nonderogatory matrices can be characterized in a number of ways, two of which are recorded in the next proposition. We postpone its proof until Proposition 3.2.4, by which time we will have collected enough ammunition to deal with it quickly.

Proposition 1.1.2

The following are equivalent for an $n \times n$ matrix A over an algebraically closed field F :

- (1) A is nonderogatory.
- (2) $\dim F[A] = n$.
- (3) The only matrices that commute with A are polynomials in A .

Nowadays, the term “nonderogatory” often goes under the name **1-regular**. The reason for this is that nonderogatory is the $k = 1$ case of a **k -regular**

4. This is equivalent to saying that left multiplying matrices by a fixed matrix A , and right multiplying matrices by a fixed matrix B , are linear transformations of $M_n(F)$.

matrix A , by which we mean a matrix whose eigenspaces are at most k -dimensional. Later in the book we will be particularly interested in 2-regular matrices and, to a lesser extent, in 3-regular matrices.

In a section on the most basic notions of linear algebra, it would be remiss of the authors not to mention *elementary row operations*, and their role in finding a basis for the null space of a matrix, for example.⁵ One should never underestimate the importance of being able to do row operations systematically, accurately, and quickly. They are the “calculus” of linear algebra. One should have the same facility with them as in differentiating and integrating elementary functions in the other Calculus.

Recall that there are three types of elementary row operations: (1) row swaps, (2) adding a multiple of one row to another (different) row, and (3) multiplying a row by a nonzero scalar. We denote the corresponding elementary matrices that produce these row operations, under left multiplication, respectively by E_{ij} (the identity matrix with rows i and j swapped), $E_{ij}(c)$ (the identity matrix with c times its row j added to row i), and $E_i(c)$ (the identity matrix with row i multiplied by the nonzero c). We will also have occasion to employ elementary *column* operations that correspond to *right* multiplication by elementary matrices. Note, however, that right multiplication by our above $E_{ij}(c)$ adds c times column i to column j , **not** column j to column i .

Here is a simple example to remind us of the computations involved in elementary row operations. In this one example, to encourage good habits, we will actually label each row operation using the lowercase version of the corresponding elementary matrix (e.g., e_{35} swaps rows 3 and 5, $e_{21}(-4)$ adds -4 times row 1 to row 2, and $e_4(\frac{2}{3})$ multiplies row 4 by $\frac{2}{3}$). We won't spell that out in later uses (in fact, later e_{ij} will be reserved for something different—the “matrix unit” having a 1 in the (i, j) position and 0's elsewhere).

Example 1.1.3

Finding a basis for the null space of the matrix

$$A = \begin{bmatrix} 1 & 2 & 0 & 2 & -1 \\ 2 & 4 & 0 & 4 & -2 \\ 1 & 2 & 3 & -1 & 8 \\ -1 & -2 & 2 & -4 & 7 \\ 3 & 6 & 0 & 6 & -3 \end{bmatrix}$$

5. Repeated use of this procedure is really the key to computing the Weyr form of a matrix, as we shall see in Chapter 2.

amounts to solving the homogeneous linear system $A\mathbf{x} = \mathbf{0}$, which in turn can be solved by putting A in (reduced) row-echelon form:

$$A \longrightarrow \begin{bmatrix} 1 & 2 & 0 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & -3 & 9 \\ 0 & 0 & 2 & -2 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{using } \begin{array}{l} e_{21}(-2), \\ e_{31}(-1), \\ e_{41}(1), \\ e_{51}(-3) \end{array}$$

$$\xrightarrow{e_{23}} \begin{bmatrix} 1 & 2 & 0 & 2 & -1 \\ 0 & 0 & 3 & -3 & 9 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{e_2(\frac{1}{3})} \begin{bmatrix} 1 & 2 & 0 & 2 & -1 \\ 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{e_{42}(-2)} \begin{bmatrix} 1 & 2 & 0 & 2 & -1 \\ 0 & 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The final matrix is the reduced row-echelon form of A , with its leading 1's in columns 1 and 3. Hence, in terms of the solution vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

of $A\mathbf{x} = \mathbf{0}$, we see that x_2, x_4, x_5 are the free variables (which can be assigned any values) and x_1, x_3 are the leading variables (whose values are determined by

the assigned free values). When we separate out the two classes of variables, the reduced row-echelon matrix gives us the equivalent linear system

$$\begin{array}{rclclcl} x_1 & = & -2x_2 & - & 2x_4 & + & x_5 \\ x_2 & = & x_2 & & & & \\ x_3 & = & & & x_4 & - & 3x_5 \\ x_4 & = & & & x_4 & & \\ x_5 & = & & & & & x_5 \end{array}$$

Recasting these equations using column vectors, we get

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 1 \\ 0 \\ -3 \\ 0 \\ 1 \end{bmatrix}.$$

Expressed another way,

$$\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for the null space of A . □

1.2 BLOCKED MATRICES

Staring at very large matrices can give one a headache, especially if the matrices require some sort of analysis under algebraic operations. So we should always be on the lookout for patterns, inductive arguments, and shortcuts. From a purely numerical analysis point of view, sparseness (lots of zeros) is often enough. But we are after something different that applies to even sparse matrices—the notion of “blocking” a matrix. It is a most useful tool. One can get by without much of an understanding of blocking in the case of the Jordan form. But the reader is warned that **an appreciation of blocked matrices is indispensable for a full understanding of our Weyr form**. There is not a lot to this. However, for whatever reason, blocking of matrices doesn’t seem to come naturally to some (even seasoned) mathematicians. Of course, every applied linear algebraist knows this stuff inside and out.⁶

6. The authors do not regard themselves as specialists in applied linear algebra.

To keep the discussion simple, we will work with square matrices A over an arbitrary field. We can **partition** the matrix A by choosing some horizontal partitioning of the rows and, independently, some vertical partitioning of the columns. For instance,

$$A = \left[\begin{array}{cc|ccc|c} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

combines the horizontal partitioning $6 = 3 + 1 + 1 + 1$ with the vertical partitioning $6 = 1 + 1 + 3 + 1$. This is a perfectly legitimate operation and very useful in some circumstances. But this particular partitioning of A is not a blocking in the sense we use the term, because if we have another 6×6 matrix B partitioned the same way, we have no additional insight into how to compute the product AB . Blocking of a matrix comes when we choose the *same* partitioning for the columns as for the rows. For instance, using the same A , we could choose the horizontal and vertical partitioning $6 = n = n_1 + n_2 + n_3 + n_4 = 2 + 2 + 1 + 1$ to give:

$$A = \left[\begin{array}{cc|cc|c|c} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] = \left[\begin{array}{c|c|c|c} A_{11} & A_{12} & A_{13} & A_{14} \\ \hline A_{21} & A_{22} & A_{23} & A_{24} \\ \hline A_{31} & A_{32} & A_{33} & A_{34} \\ \hline A_{41} & A_{42} & A_{43} & A_{44} \end{array} \right] = (A_{ij}),$$

where the A_{ij} are the $n_i \times n_j$ submatrices given by the rectangular partitioning. For example,

$$A_{12} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, A_{23} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, A_{34} = \begin{bmatrix} 1 \end{bmatrix}.$$

In this context, where the same partition is used for both the rows and the columns, A is referred to as a **block** or **blocked matrix** and each A_{ij} as its (i, j) th **block**. Note that the diagonal blocks A_{ii} are all square submatrices.

Now given another 6×6 matrix B blocked in the same way (using the same partition), there is additional insight into how to compute the product AB . For instance, if

$$B = \left[\begin{array}{cc|cc|c|c} 1 & 2 & 3 & 1 & 1 & 5 \\ 3 & 4 & 1 & 1 & 0 & 2 \\ \hline 7 & 1 & 3 & 3 & 2 & 1 \\ 1 & 1 & 8 & 2 & 3 & 4 \\ \hline 9 & 5 & 6 & 1 & 2 & 7 \\ 6 & 0 & 1 & 8 & 1 & 2 \end{array} \right] = \left[\begin{array}{c|c|c|c} B_{11} & B_{12} & B_{13} & B_{14} \\ \hline B_{21} & B_{22} & B_{23} & B_{24} \\ \hline B_{31} & B_{32} & B_{33} & B_{34} \\ \hline B_{41} & B_{42} & B_{43} & B_{44} \end{array} \right] = (B_{ij}),$$

then

$$AB = \left[\begin{array}{cc|cc|c|c} 7 & 1 & 3 & 3 & 2 & 1 \\ 1 & 1 & 8 & 2 & 3 & 4 \\ \hline 9 & 5 & 6 & 1 & 2 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 6 & 0 & 1 & 8 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right].$$

Of course, one can get this by multiplying the 6×6 matrices in the usual way. Or one can multiply the pair of blocked matrices $A = (A_{ij})$, $B = (B_{ij})$, by the usual rule of matrix multiplication for 4×4 matrices, but viewing the entries of the new matrices as themselves matrices (the A_{ij} , B_{ij}) of various sizes. Since we have partitioned the rows and columns the same way, the internal matrix calculations for the product will involve matrices of compatible size. For instance, the $(1, 2)$ block entry of each of the blocked matrices is an ordinary 2×2 matrix. In the product AB , the $(1, 2)$ block entry becomes

$$\begin{aligned} \sum_{k=1}^4 A_{1k}B_{k2} &= A_{11}B_{12} + A_{12}B_{22} + A_{13}B_{32} + A_{14}B_{42} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 3 \\ 8 & 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 6 & 1 \end{bmatrix} \\ &\quad + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 8 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 3 \\ 8 & 2 \end{bmatrix}. \end{aligned}$$

Having done this for our particular B , one can spot the pattern in AB for any B , for this fixed A . (What is it?) But it requires the blocked matrix view to see this pattern in its clearest form.

Of course, one can justify the multiplication of blocked matrices in general (those sharing the same blocking), without getting into a subscript frenzy. Our reader can look at the Horn and Johnson text *Matrix Analysis*, or the article by Reams in the *Handbook of Linear Algebra*, for more general discussions on matrix partitioning.

Notice that in specifying the block structure of a blocked matrix $A = (A_{ij})$, we need only specify the sizes of the (square) diagonal blocks A_{ii} , because the (i, j) block A_{ij} must be $n_i \times n_j$ where n_i and n_j are the i th and j th diagonal block sizes, respectively. Moreover, as will nearly always be the case with our blocked matrices, if the diagonal blocks have decreasing size, the whole block structure of an $n \times n$ matrix can be specified uniquely simply by a partition $n_1 + n_2 + \cdots + n_r = n$ of n with $n_1 \geq n_2 \geq \cdots \geq n_r \geq 1$. The simplest picture occurs when $n_1 = n_2 = \cdots = n_r = d$, because blocking an $n \times n$ matrix this way just amounts to viewing it as an $r \times r$ matrix over the ring $M_d(F)$ of $d \times d$ matrices.

If $A = (A_{ij})$ is a blocked matrix in which the $A_{ij} = 0$ for $i > j$, that is, all the blocks below the diagonal are zero, then A is said to be **block upper triangular**. It should be clear to the reader what we mean by **strictly block upper triangular** and **(strictly) block lower triangular**.

Our example A above is block upper triangular. We can (and will) simplify the picture for a block upper triangular matrix by leaving the lower (zero) blocks blank, so that, for our example we have

$$A = \left[\begin{array}{cc|cc|cc} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline & & 0 & 0 & 1 & 0 \\ & & 0 & 0 & 0 & 0 \\ \hline & & & & 0 & 1 \\ & & & & & 0 \end{array} \right].$$

The reader may not know it (nor is expected to at this stage), but she or he is looking at the 6×6 nilpotent Weyr matrix of Weyr structure $(2, 2, 1, 1)$. The point we wish to make is that our first partitioning of the same matrix is not as revealing as this blocked form.

Just as with ordinary matrices, the simplest blocked matrices A are the **block diagonal** matrices—all the off-diagonal blocks are zero:

$$A = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_r \end{bmatrix}.$$

In this case, we write $A = \text{diag}(A_1, A_2, \dots, A_r)$ and say A is a **direct sum of the matrices** A_1, A_2, \dots, A_r . If $B = \text{diag}(B_1, B_2, \dots, B_r)$ is a second block diagonal matrix (for the same blocking), then

$$AB = \text{diag}(A_1B_1, A_2B_2, \dots, A_rB_r).$$

Of course, sums and scalar multiples behave similarly, so our knowledge of a block diagonal matrix is as good as our knowledge of its individual diagonal blocks. This is a simple but fundamental observation, used again and again in canonical forms, for instance. Those with a ring theory background may prefer to view this as saying the following. For matrices blocked according to a fixed partition $n = n_1 + n_2 + \dots + n_r$, the mapping

$$\theta : (A_1, A_2, \dots, A_r) \longmapsto \text{diag}(A_1, A_2, \dots, A_r)$$

is an algebra isomorphism (1-1 correspondence preserving addition, multiplication, and scalar multiplication) of the direct product $\prod_{i=1}^r M_{n_i}(F)$ of the matrix algebras $M_{n_i}(F)$ onto the algebra of $n \times n$ block diagonal matrices (with the specified blocking).

We finish our discussion of blocked matrices with another seemingly trivial, but very useful, observation on block upper triangular matrices. The yet-to-be-described Weyr form (when in company with some other commuting matrices) is particularly amenable to this result, more so than the Jordan form. We state the result for 2×2 block upper triangular matrices, but there is an obvious extension to general block upper triangular ones.

Proposition 1.2.1

Let m and n be positive integers with $m < n$. Let \mathcal{T} be the algebra of all $n \times n$ matrices A that are block upper triangular with respect to the partition

$$n = m + (n - m):$$

$$A = \left[\begin{array}{c|c} P & Q \\ \hline 0 & R \end{array} \right],$$

where P is $m \times m$, Q is $m \times (n - m)$, and R is $(n - m) \times (n - m)$. Then the projection

$$\eta : \mathcal{T} \longrightarrow M_m(F), \quad A \longmapsto P$$

onto the top left corner is an algebra homomorphism (that is, preserves addition, multiplication, and scalar multiplication) of \mathcal{T} onto the algebra $M_m(F)$ of $m \times m$ matrices.

Proof

Clearly η preserves addition and scalar multiplication. Now let

$$A = \left[\begin{array}{c|c} P & Q \\ \hline 0 & R \end{array} \right], \quad A' = \left[\begin{array}{c|c} P' & Q' \\ \hline 0 & R' \end{array} \right]$$

be in \mathcal{T} . Since

$$AA' = \left[\begin{array}{c|c} PP' & PQ' + QR' \\ \hline 0 & RR' \end{array} \right],$$

we have $\eta(AA') = PP' = \eta(A)\eta(A')$. Thus, η is an algebra homomorphism. \square

Remarks 1.2.2

- (1) Projecting onto the bottom right corner is also a homomorphism.
- (2) Also, if \mathcal{T} is the algebra of block upper triangular matrices relative to the partition $n = n_1 + n_2 + \cdots + n_r$, then, for $1 \leq i \leq r$, the projection onto the top left-hand $i \times i$ corner of blocks is an algebra homomorphism onto the algebra of block upper triangular matrices of size $m = n_1 + n_2 + \cdots + n_i$ (relative to the implied truncated partition of m). This homomorphism is just the restriction of η in the proposition for the case $m = n_1 + n_2 + \cdots + n_i$. \square

1.3 CHANGE OF BASIS AND SIMILARITY

Change of basis and similarity are really about reformulating a given linear algebra problem into an equivalent one that is easier to tackle. (It is a bit like using equivalent frames of reference in the theory of relativity.) These fundamental processes are reversible, so if we are able to answer the simpler question, we can return with a solution to the initial problem.

Fix an n -dimensional vector space V and an (ordered) basis $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ for V . The **co-ordinate vector** of $v \in V$ relative to \mathcal{B} is

$$[v]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix},$$

where the a_i are the unique scalars for which $v = a_1 v_1 + a_2 v_2 + \dots + a_n v_n$. If \mathcal{B}' is another basis, we let $[\mathcal{B}', \mathcal{B}]$ denote the **change of basis matrix**, that is the $n \times n$ matrix whose columns are the co-ordinate vectors of the \mathcal{B}' basis vectors relative to \mathcal{B} . This is an invertible matrix with $[\mathcal{B}', \mathcal{B}]^{-1} = [\mathcal{B}, \mathcal{B}']$. Co-ordinate vectors now change according to the rule⁷

$$[v]_{\mathcal{B}'} = [\mathcal{B}, \mathcal{B}'] [v]_{\mathcal{B}}.$$

Now let $T : V \rightarrow V$ be a linear transformation. Its **matrix** $[T]_{\mathcal{B}}$ **relative to the (ordered) basis \mathcal{B}** is defined as the $n \times n$ matrix whose columns are the co-ordinate vectors $[T(v_1)]_{\mathcal{B}}, [T(v_2)]_{\mathcal{B}}, \dots, [T(v_n)]_{\mathcal{B}}$ of the images of the \mathcal{B} basis vectors. The reason why we work with columns⁸ rather than rows here is that our transformations act on the left of vectors, and our composition of two transformations is in accordance with this: $(ST)(v) = S(T(v))$. The correspondence $v \mapsto [v]_{\mathcal{B}}$ is a vector space isomorphism from V to n -space F^n . What T is doing, under this identification, is simply left multiplying column vectors by the matrix $[T]_{\mathcal{B}}$:

$$[T(v)]_{\mathcal{B}} = [T]_{\mathcal{B}} [v]_{\mathcal{B}}.$$

(So, abstractly, a linear transformation is just left multiplication of column vectors by a matrix.)

7. A good way to remember this and other change of basis results is that primed and unprimed basis labels alternate.

8. This is the sensible rule, but unfortunately not all authors observe it. To break it invites trouble in the more general setting of vector spaces over division rings. (In that setting, one should also place the scalars on the right of vectors.)

A **permutation matrix** is a square $n \times n$ matrix P whose rows (resp. columns) are a permutation of the rows (resp. columns) of the identity matrix I under some permutation $p \in S_n$ (resp. $p^{-1} \in S_n$). (Here, S_n is the *symmetric group* of all permutations of $1, 2, \dots, n$.) In terms of the matrix of a linear transformation, and in the case of a row permutation p , we have $P = [T]_{\mathcal{B}}$ where $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ is the standard basis of F^n and $T : F^n \rightarrow F^n$ is the linear transformation whose action on \mathcal{B} is $T(v_i) = v_{p(i)}$.

For fixed V and basis \mathcal{B} , the correspondence $T \mapsto [T]_{\mathcal{B}}$ provides the fundamental isomorphism between the algebra $\mathcal{L}(V)$ of all linear transformations of V (to itself) and the algebra $M_n(F)$ of all $n \times n$ matrices over F : it is a 1-1 correspondence that preserves sums, products⁹ and scalar multiples. The result should be etched in the mind of every serious student of linear algebra.¹⁰

Two square $n \times n$ matrices A and B are called **similar** if $B = C^{-1}AC$ for some invertible matrix C . “Similar” is an understatement here, because A and B will have identical algebraic properties. (In particular, similar matrices have the same eigenvalues, determinant, rank, trace,¹¹ and so on.) This is because for a fixed invertible C , and a variable matrix A , the conjugation mapping $A \mapsto C^{-1}AC$ is an algebra automorphism of $M_n(F)$ (a 1-1 correspondence preserving sums, products, and scalar multiples).¹² And under an automorphism (or isomorphism), an element and its image have the same algebraic properties.¹³ This view of similarity is entirely analogous to, for example, conjugation in group theory. But what is new in the linear algebra setting is how nicely similarity relates to the matrices of a linear transformation $T : V \rightarrow V$ of an n -dimensional

9. If we had put the co-ordinate vectors $[T(v_i)]_{\mathcal{B}}$ as rows of the representing matrix, the correspondence would reverse products.

10. Unfortunately, nowadays some otherwise very good students come away from linear algebra courses without ever having seen this.

11. The trace, $\text{tr } A$, of a square matrix A is the sum of its diagonal entries.

12. The so-called Skolem–Noether theorem of ring theory tells us that these conjugations are the only algebra automorphisms of $M_n(F)$. (See Jacobson’s *Basic Algebra II*, p. 222.)

13. Thinking of complex conjugation as an automorphism of \mathbb{C} , we see that a complex number and its conjugate are algebraically indistinguishable. In particular, there is really no such thing as “the” (natural) complex number i satisfying $i^2 = -1$, short of arbitrarily nominating one of the two roots (because the two solutions are conjugates). This is unlike the distinction between, say, the two square roots of 2 in \mathbb{R} . Here, one root is positive, hence expressible as a square of a real number; the other is not. So the two can be distinguished by an algebraic property.

space under a change of basis from \mathcal{B} to \mathcal{B}' . They are always similar:

$$[T]_{\mathcal{B}'} = C^{-1} [T]_{\mathcal{B}} C$$

where $C = [\mathcal{B}', \mathcal{B}]$.

Moreover, every pair of similar matrices can be viewed as the matrices of a single transformation relative to suitable bases.

A useful observation in the case that C is a permutation matrix, corresponding to some permutation $p \in S_n$, but this time via the action of p on the columns of I , is that $C^{-1}AC$ is the matrix obtained by first permuting the *columns* of A under p , and then permuting the *rows* of the resulting matrix by the *same* permutation p . For instance, if $p = (1\ 2\ 3)$ is the cyclic permutation, then

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad C^{-1} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} C = \begin{bmatrix} i & g & h \\ c & a & b \\ f & d & e \end{bmatrix}.$$

A standard way of utilizing the transformation view of similarity, but with a matrix outcome in mind, is this: suppose we are presented with an $n \times n$ matrix A over the field F and we are looking for a simpler matrix B (perhaps diagonal) to which A is similar. Firstly, let $V = F^n$, let \mathcal{B} be the standard basis for V , and let $T : V \rightarrow V$ be the linear transformation that left multiplies column vectors by A :

$$T \left(\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \right) = A \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Note $[T]_{\mathcal{B}} = A$. Secondly, “using one’s wits” (depending on additional information about A), find another basis \mathcal{B}' relative to which the matrix $B = [T]_{\mathcal{B}'}$ looks nice. Thirdly, let $C = [\mathcal{B}', \mathcal{B}]$ be the change of basis matrix. Note that C has the \mathcal{B}' basis vectors as its columns and is invertible. Now we have our similarity $B = C^{-1}AC$ by the change of basis result for the matrices of a transformation.

Again, suppose $T : V \rightarrow V$ is a linear transformation of an n -dimensional space. A subspace U of V is said to be **invariant** under T if $T(U) \subseteq U$ (T maps vectors of U into U). Notice that a nonzero vector $v \in V$ is an eigenvector of T (for some eigenvalue) precisely when $\langle v \rangle$ is invariant under T . (This provides a clear geometric picture of why a proper rotation of the real plane

about the origin through less than 180 degrees can't have any eigenvalues—no lines through the origin are invariant under the rotation.) If we choose a basis \mathcal{B}_1 for an invariant subspace U and extend it to a basis \mathcal{B} of V , then the matrix of T relative to \mathcal{B} is block upper triangular in which the top left block is $m \times m$, where $m = \dim U$, and the bottom right block is $(n - m) \times (n - m)$:

$$[T]_{\mathcal{B}} = \left[\begin{array}{c|c} P & Q \\ \hline 0 & R \end{array} \right],$$

where P is the matrix of $T|_U : U \rightarrow U$ relative to \mathcal{B}_1 . This observation can often be used as an inductive tool. (It also allows a neat noninductive proof of the Cayley–Hamilton theorem in terms of transformations, by fixing $v \in V$ and taking U to be the subspace spanned by all the $T^i(v)$. Then through a natural choice for \mathcal{B}_1 , the matrix P is a “companion matrix” whose characteristic polynomial $p(x) \in M_m(F)$ is easily calculated, and for which $p(T)(v) = 0$ is easily verified. The reader is invited to complete the argument, or to curse the authors for not doing so!) The kernel and image of a transformation T are always subspaces invariant under T . We record the following simple generalization.

Proposition 1.3.1

Suppose S and T are commuting linear transformations of a vector space V . Then the kernel and image of S are subspaces which are invariant under T .

Proof

Let $U = \ker S$. For $u \in U$, we have

$$\begin{aligned} S(T(u)) &= (ST)(u) \\ &= (TS)(u) \quad (\text{by commutativity}) \\ &= T(S(u)) \\ &= T(0) \\ &= 0, \end{aligned}$$

which shows $T(u) \in U$. Thus, U is invariant under T . Similarly, so is $S(V)$. \square

A vector space V is a **direct sum** of subspaces U_1, U_2, \dots, U_k , written

$$V = U_1 \oplus U_2 \oplus \cdots \oplus U_k,$$

if every $v \in V$ can be written uniquely as $v = u_1 + u_2 + \cdots + u_k$, where each $u_i \in U_i$. In this case, a union of linearly independent subsets from each of the U_i

remains linearly independent. Consequently, $\dim V = \dim U_1 + \dim U_2 + \cdots + \dim U_k$. As with (internal) direct sums or products of other algebraic structures, one can verify that a sum $U_1 + U_2 + \cdots + U_k$ of subspaces is a direct sum, meaning $U_1 + U_2 + \cdots + U_k = U_1 \oplus U_2 \oplus \cdots \oplus U_k$, by repeated use of the condition that for $k = 2$, directness means that $U_1 \cap U_2 = 0$. In general, we check the “triangular conditions”:

$$\begin{aligned} U_1 \cap U_2 &= 0, \\ (U_1 + U_2) \cap U_3 &= 0, \\ (U_1 + U_2 + U_3) \cap U_4 &= 0, \\ &\vdots \\ (U_1 + U_2 + U_3 + \cdots + U_{k-1}) \cap U_k &= 0. \end{aligned}$$

An especially useful observation (when teamed with results for change of basis matrices) is the following.

Proposition 1.3.2

Suppose $T : V \rightarrow V$ is a linear transformation and

$$V = U_1 \oplus U_2 \oplus \cdots \oplus U_k$$

is a direct sum decomposition of V into T -invariant subspaces U_1, U_2, \dots, U_k . Pick a basis \mathcal{B}_i for each U_i and let $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$. Then relative to the basis \mathcal{B} for V , the matrix of T is the block diagonal matrix

$$[T]_{\mathcal{B}} = \begin{bmatrix} \boxed{A_1} & & & \\ & \boxed{A_2} & & \\ & & \ddots & \\ & & & \boxed{A_k} \end{bmatrix},$$

where A_i is the matrix relative to \mathcal{B}_i of the restriction of T to U_i .

Proof

There is nothing to this if (1) we have a clear mental picture of what the matrix of a transformation relative to a specified basis looks like,¹⁴ and (2) appreciate that the restriction of a linear transformation T to an invariant subspace U is a

14. If one is constantly referring back to the definition of the matrix of a transformation, and consulting with the “subscript doctor,” this distraction may hamper progress in later chapters.

linear transformation of U as a vector space in its own right. For instance, suppose $\dim U_1 = 3$ and $\dim U_2 = 2$ and we label the basis vectors by $\mathcal{B}_1 = \{v_1, v_2, v_3\}$ and $\mathcal{B}_2 = \{v_4, v_5\}$. Since U_1 is T -invariant, for $i = 1, 2, 3$, the $T(v_i)$ are linear combinations of only v_1, v_2, v_3 , so in the matrix $[T]_{\mathcal{B}}$, the first three columns have zeros past row three. Similarly, for $i = 4, 5$, the $T(v_i)$ are linear combinations of only v_4, v_5 , so in the matrix $[T]_{\mathcal{B}}$, columns four and five have no nonzero entries outside of rows four and five. And so on. \square

1.4 DIAGONALIZATION

There isn't a question that one can't immediately answer about a diagonal matrix

$$D = \text{diag}(d_1, d_2, \dots, d_n) = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & d_n \end{bmatrix}.$$

For instance, its k th power is $\text{diag}(d_1^k, d_2^k, \dots, d_n^k)$. So it is of interest to know when a square $n \times n$ matrix A is similar to a diagonal matrix. (Then, for example, its powers can also be computed.) Such a matrix A is called **diagonalizable**: there exists an invertible matrix C such that $C^{-1}AC$ is diagonal. Standard texts include many interesting applications of diagonalizable matrices, from Markov processes, to finding principal axes of quadratic forms, through to solving systems of first order linear differential equations. Later, in Chapter 6, we examine an "approximate" version of diagonalization, which has modern relevance to phylogenetics and multivariate interpolation.

Conceptually, the key to understanding diagonalization is through linear transformations $T : V \rightarrow V$. The matrix of T relative to a basis \mathcal{B} is diagonal precisely when the basis vectors are eigenvectors for various eigenvalues. In that case, the matrix is simply the diagonal matrix of the matching eigenvalues, in the order the basis vectors happen to be presented. An individual eigenvalue will appear on the diagonal according to its algebraic multiplicity. (This is just Proposition 1.3.2 when all the U_i are one-dimensional.) Sensibly, one should reorder the basis vectors to group together those sharing the same eigenvalue. To connect all this with matrices, we just use change of basis results.

Presented with a small $n \times n$ matrix A , whose eigenvalues we know, we can test if A is diagonalizable by checking if the geometric multiplicities of its various eigenvalues sum to n . And an explicit C that diagonalizes A can also be found. Here is an example to remind us of the process.

Example 1.4.1

Suppose we wish to diagonalize the real matrix

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}.$$

Using a first row cofactor expansion, we see that the characteristic polynomial $p(x)$ of A is

$$\begin{aligned} p(x) &= \det \begin{bmatrix} x-3 & -1 & -1 \\ -1 & x-3 & -1 \\ -1 & -1 & x-3 \end{bmatrix} \\ &= (x-3)[(x-3)^2 - 1] + 1(-x+3-1) - 1(1+x-3) \\ &= (x-2)^2(x-5). \end{aligned}$$

Hence the eigenvalues of A are 2, 5 with respective algebraic multiplicities 2 and 1. We need to check if these agree with the geometric multiplicities. We can compute a basis for the eigenspace $E(2)$ using elementary row operations:

$$2I - A = \begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The corresponding homogeneous system $(2I - A)\mathbf{x} = 0$ has two free variables, from which we can pick out the basis

$$\mathcal{B}_2 = \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

for $E(2)$. For $E(5)$ we proceed similarly:

$$\begin{aligned} 5I - A &= \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & -2 \\ -1 & 2 & -1 \\ 2 & -1 & -1 \end{bmatrix} \\ &\longrightarrow \begin{bmatrix} 1 & 1 & -2 \\ 0 & 3 & -3 \\ 0 & -3 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & -2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

This gives one free variable in the homogeneous system $(5I - A)\mathbf{x} = 0$, from which we get the basis

$$\mathcal{B}_5 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

for the eigenspace $E(5)$. Thus, the geometric multiplicities of the eigenvalues 2 and 5 sum to $n = 3$, whence A is diagonalizable.

We can diagonalize A explicitly with an invertible matrix C as follows. Let \mathcal{B} be the standard basis for $V = F^3$ and note A is the matrix of its left multiplication map of V relative to \mathcal{B} . Next, form the basis for V

$$\mathcal{B}' = \mathcal{B}_2 \cup \mathcal{B}_5 = \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

of eigenvectors of A . Finally take these basis vectors as the columns of the matrix

$$C = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

The outcome, by the change of basis result for a linear transformation (looking at the left multiplication map by A relative to \mathcal{B}' and noting that $C = [\mathcal{B}', \mathcal{B}]$), is

$$C^{-1}AC = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix},$$

a diagonal matrix having the eigenvalues 2 and 5 on the diagonal and repeated according to their algebraic multiplicities. \square

In the above example, the diagonalization works over any field F whose characteristic is not 3. (When F has characteristic 3, the above A has 2 as its only eigenvalue but this has geometric multiplicity only 2, less than 3.) So, in general, diagonalization depends on the base field. For instance, real symmetric and complex hermitian matrices are always diagonalizable (in fact by an orthogonal and unitary matrix, respectively), but over the two element field, only the idempotent matrices E (those satisfying $E^2 = E$) are diagonalizable. A frequently used observation is that an $n \times n$ matrix that has n distinct eigenvalues is diagonalizable. The general theorem is that $A \in M_n(F)$ is diagonalizable if and only if the minimal polynomial of A factors into distinct linear factors. The **minimal polynomial** is the unique monic polynomial $m(x)$

of least degree such that $m(A) = 0$. It can be calculated by finding the first power A^s of A that is linearly dependent on the earlier powers $I, A, A^2, \dots, A^{s-1}$, say $A^s = c_0I + c_1A + \dots + c_{s-1}A^{s-1}$, and taking

$$m(x) = x^s - c_{s-1}x^{s-1} - \dots - c_1x - c_0.$$

The minimal polynomial divides all other polynomials that vanish at A . In particular, by the Cayley–Hamilton theorem, the minimal polynomial divides the characteristic polynomial, so the degree of $m(x)$ is at most n . In fact, $m(x)$ has the same zeros as the characteristic polynomial (the eigenvalues of A), only with smaller multiplicities. In some ways, the minimal polynomial is more revealing of the properties of a matrix than the characteristic polynomial. One can also show that, as an ideal of $F[x]$, the kernel of the polynomial evaluation map $f \mapsto f(A)$ has the minimal polynomial of A as its monic generator. This is as good a place as any to record another property of the minimal polynomial, which we use (sometimes implicitly) in later chapters.

Proposition 1.4.2

The dimension of the subalgebra $F[A]$ generated by a square matrix $A \in M_n(F)$ agrees with the degree of the minimal polynomial $m(x)$ of A .

Proof

Finite-dimensionality of $M_n(F)$ guarantees some power of A is dependent on earlier powers, so there is a least such power A^s that is so dependent. Let

$$(*) \quad A^s = c_0I + c_1A + \dots + c_{s-1}A^{s-1}$$

be the corresponding dependence relation. Now $I, A, A^2, \dots, A^{s-1}$ all lie in $F[A]$ and are linearly independent by choice of s . We need only show they span $F[A]$ in order to conclude they form a basis with s members, whence $\dim F[A] = s = \deg(m(x))$. In turn, since the powers of A span $F[A]$, it is enough to get these powers as linear combinations of $I, A, A^2, \dots, A^{s-1}$. But this just involves repeated applications of the relationship $(*)$:

$$\begin{aligned} A^{s+1} &= AA^s \\ &= A(c_0I + c_1A + \dots + c_{s-1}A^{s-1}) \\ &= c_0A + c_1A^2 + \dots + c_{s-1}A^s \\ &= c_0A + c_1A^2 + \dots + c_{s-2}A^{s-1} + c_{s-1}(c_0I + c_1A + \dots + c_{s-1}A^{s-1}) \\ &= c_{s-1}c_0I + (c_0 + c_{s-1}c_1)A + \dots + (c_{s-2} + c_{s-1}^2)A^{s-1} \end{aligned}$$

and so on. □

Unlike the minimal polynomial of an algebraic field element, minimal polynomials of matrices need not be irreducible. In fact, *any* monic polynomial of positive degree can be the minimal polynomial of a suitable matrix, and the same can happen for the characteristic polynomial.¹⁵ The following is the standard example:

Example 1.4.3

Let $f(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_2x^2 + c_1x + c_0 \in F[x]$ be a monic polynomial of degree n . Then the following “companion matrix”

$$C = \begin{bmatrix} 0 & 0 & \cdots & & -c_0 \\ 1 & 0 & & & -c_1 \\ 0 & 1 & & & -c_2 \\ \vdots & & \ddots & & \vdots \\ & & & 1 & 0 & -c_{n-2} \\ 0 & 0 & \cdots & 0 & 1 & -c_{n-1} \end{bmatrix}$$

has $f(x)$ as both its minimal and characteristic polynomials. To see this, one observes that the first n powers of C are independent so the minimal polynomial has degree n and necessarily agrees with the characteristic polynomial. On the other hand, the characteristic polynomial $\det(xI - C)$ can easily be computed to be $f(x)$ directly, by a cofactor expansion in the first row (combined with induction for matrices of size $n - 1$ when evaluating the $(1, 1)$ cofactor). \square

Another useful point of view of diagonalizable matrices $A \in M_n(F)$ is that they are precisely the matrices possessing a “spectral resolution”:

$$A = \lambda_1 E_1 + \lambda_2 E_2 + \cdots + \lambda_k E_k$$

where the λ_i are scalars and the E_i are orthogonal idempotent matrices, that is, $E_i^2 = E_i$ and $E_i E_j = 0$ for $i \neq j$. (In the spectral resolution, there is no loss of generality in assuming that the λ_i are distinct, in which case the E_i are actually polynomials in A .) This is a nice “basis-free” approach.

15. Consequently, there can be no way of exactly computing the eigenvalues of a general matrix. Nor can there be a “formula” for the eigenvalues in terms of the rational operations of addition, multiplication, division, and extraction of m th roots on the entries of a general matrix of size bigger than 4×4 . (This follows from Galois theory, more particularly the Abel–Ruffini theorem that quintic and higher degree polynomial equations are not “solvable by radicals.”) However, the matrices that arise in practice (e.g., tridiagonal) are often amenable to fast, high-precision, eigenvalue methods.

Generally, idempotent matrices play an important role in matrix theory. To within similarity, an idempotent matrix E looks like the diagonal matrix

$$E = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

where the number of 1's is the rank of E . In particular, idempotent matrices have only 0 and 1 as eigenvalues (combined with diagonalizability, this characterizes idempotents). Idempotent linear transformations $T : V \rightarrow V$ are exactly the projection maps

$$T : U \oplus W \longrightarrow U, u + w \longmapsto u$$

associated with direct sum decompositions $V = U \oplus W$. Necessarily, U is the image of T , on which T acts as the identity transformation, and W is its kernel, on which T acts, of course, as the zero transformation. Since both these subspaces are T -invariant, a quick application of Proposition 1.3.2 gives the displayed idempotent matrix E of T for a suitable basis. In turn, by change of basis results, this justifies the above claim that idempotent matrices look just like E to within similarity. Again it is a transformation view that has led us to a nice matrix conclusion.

1.5 THE GENERALIZED EIGENSPACE DECOMPOSITION

Recall that a matrix N is nilpotent if $N^r = 0$ for some positive integer r , and the least such r is called the nilpotency index of N . When our base field F is algebraically closed, many problems in linear algebra reduce to the case of nilpotent matrices. In particular, this is true in establishing the Jordan and Weyr canonical forms. The reduction is best achieved through the generalized eigenspace decomposition, which we will describe in this section.

Nice though they are, diagonalizable matrices, at least those occurring in practice, form only a small class of matrices.¹⁶ The analysis of a general matrix requires a canonical form such as the rational, Jordan, or Weyr form, which

16. The relative size of the class of diagonalizable matrices in $M_n(F)$ depends, of course, on the base field F and the order n of the matrices. For example, when $n = 2$ and F is the two element field, 8 out of the 16 matrices are diagonalizable. Things don't improve in $M_2(\mathbb{R})$. Here a randomly chosen matrix still has only a 50% chance of being diagonalizable. Of course in $M_n(\mathbb{C})$ for any n , with probability 1 a randomly chosen matrix will be diagonalizable because its eigenvalues will be distinct. However, when the eigenvalues are known not to be distinct, the

encompass more than just diagonal matrices. When the field F is algebraically closed, a general matrix $A \in M_n(F)$ is the sum of a diagonalizable matrix D and a nilpotent matrix N , which commute. This follows quickly from the generalized eigenspace decomposition. Therefore, since diagonalizable matrices are those possessing a spectral resolution, in a sense, all matrices are put together in terms of idempotent, nilpotent, and scalar matrices. We already understand the structure of the diagonalizable part D . Understanding the nilpotent part N is more involved, but the Jordan and Weyr forms describe it completely, as we will see in Section 1.7 and in Chapter 2.

Fix $A \in M_n(F)$. The **generalized eigenspace** of A corresponding to an eigenvalue λ of A is

$$G(\lambda) = \{\mathbf{x} \in F^n : (\lambda I - A)^m \mathbf{x} = 0 \text{ for some } m \in \mathbb{N}\}.$$

Clearly, $G(\lambda) \supseteq E(\lambda)$. Suppose F is algebraically closed. Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be the distinct eigenvalues of A and let

$$p(x) = (x - \lambda_1)^{m_1} (x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}$$

be the factorization of the characteristic polynomial of A into linear factors. Then one can show that

$$\begin{aligned} G(\lambda_i) &= \{\mathbf{x} \in F^n : (\lambda_i I - A)^{m_i} \mathbf{x} = 0\} \\ &= \ker(\lambda_i I - A)^{m_i}. \end{aligned}$$

The first description of a generalized eigenspace has the advantage of not referencing the characteristic polynomial. However, the reader may prefer to take this second description of $G(\lambda_i)$, in terms of the algebraic multiplicity m_i of λ_i , as the definition of a generalized eigenspace. That saves proving it agrees with the first, which strictly speaking involves establishing part of the generalized eigenspace decomposition 1.5.2.

An $n \times n$ diagonalizable matrix A is characterized by the property that n -space F^n is a direct sum of the eigenspaces of A :

$$F^n = E(\lambda_1) \oplus E(\lambda_2) \oplus \cdots \oplus E(\lambda_k),$$

where $\lambda_1, \dots, \lambda_k$ are the distinct eigenvalues of A . (The sum itself is always direct for any matrix, but may not fill F^n .) Using generalized eigenspaces, we

odds drop quite dramatically. For instance, the probability of diagonalizability of a 6×6 complex matrix whose eigenvalues are, say, $-1, 2, 2, 2, 4, 4$ is less than $1/6$. And at the extreme end, the chance of a random complex $n \times n$ matrix with no distinct eigenvalues being diagonalizable when $n \geq 2$ is 0.

get a direct sum decomposition for all matrices. This follows as a corollary to the so-called primary decomposition theorem given below, which works over an arbitrary field. Since the generalized eigenspace decomposition is not always readily accessible in undergraduate linear algebra texts, we will outline the proofs.

Theorem 1.5.1 (Primary Decomposition Theorem)

Suppose $T : V \rightarrow V$ is a linear transformation of an n -dimensional space V over a field F , and $p(T) = 0$ for some nonconstant monic polynomial $p(x) \in F[x]$. Let

$$p = p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$$

be the factorization of p into monic irreducible polynomials where p_1, p_2, \dots, p_k are distinct. For $i = 1, 2, \dots, k$, let $W_i = \ker p_i(T)^{m_i}$. Then the subspaces W_i are invariant under T , and V is their direct sum :

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_k.$$

Proof

In essence, the proof is the same one used to establish the primary decomposition of a finite abelian group into p -groups. By Proposition 1.3.1, the W_i are T -invariant. For each i , let $f_i = p/p_i^{m_i}$. Since these polynomials are relatively prime, there exist polynomials g_1, g_2, \dots, g_k such that

$$f_1 g_1 + f_2 g_2 + \cdots + f_k g_k = 1.$$

It follows that $f_1(T)g_1(T) + f_2(T)g_2(T) + \cdots + f_k(T)g_k(T) = I$. Now given $v \in V$, we have $v = I(v) = \sum_{i=1}^k f_i(T)g_i(T)(v)$ with $f_i(T)g_i(T)(v) \in W_i$ because $p_i(T)^{m_i} f_i(T)g_i(T)(v) = p(T)g_i(T)(v) = 0$. Thus,

$$V = W_1 + W_2 + \cdots + W_k.$$

To show this sum is direct, suppose $w_1 + w_2 + \cdots + w_k = 0$ for some $w_i \in W_i$. We need to show that all $w_i = 0$. Fix i . Observe that f_i and $p_i^{m_i}$ are relatively prime polynomials, so there are polynomials $q_i, r_i \in F[x]$ with $q_i f_i + r_i p_i^{m_i} = 1$. Now, after noting that $p_i^{m_i}(T)(w_i) = 0$ and $f_j(T)(w_j) = 0$ for $j \neq i$, we have

$$\begin{aligned} w_i &= I(w_i) \\ &= (q_i(T)f_i(T) + r_i(T)p_i^{m_i}(T))(w_i) \\ &= q_i(T)f_i(T)(w_i) \\ &= q_i(T)f_i(T)(w_1 + w_2 + \cdots + w_k) \end{aligned}$$

$$\begin{aligned}
&= q_i(T)f_i(T)(0) \\
&= 0,
\end{aligned}$$

which completes the proof. \square

Theorem 1.5.2 (The Generalized Eigenspace Decomposition)

For an $n \times n$ matrix A over an algebraically closed field F , the space F^n is the direct sum of the generalized eigenspaces of A :

$$F^n = G(\lambda_1) \oplus G(\lambda_2) \oplus \cdots \oplus G(\lambda_k),$$

where $\lambda_1, \dots, \lambda_k$ are the distinct eigenvalues of A .

Proof

Let $V = F^n$, take T to be the left multiplication map by A , and p the characteristic polynomial. By the Cayley–Hamilton theorem, $p(T)(V) = p(A)(V) = 0$ and hence $p(T) = 0$. The W_i in the primary decomposition theorem are now just the generalized eigenspaces. \square

Example 1.5.3

Let's illustrate the generalized eigenspace decomposition with the following simple example:

$$A = \begin{bmatrix} 4 & 1 & -1 \\ 0 & 2 & 2 \\ -1 & -1 & 4 \end{bmatrix}.$$

The characteristic polynomial of A is

$$p(x) = \det(xI - A) = \det \begin{bmatrix} x-4 & -1 & 1 \\ 0 & x-2 & -2 \\ 1 & 1 & x-4 \end{bmatrix} = (x-3)^2(x-4)$$

so A has eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 4$ with respective algebraic multiplicities 2 and 1.

We compute a basis for the first generalized eigenspace $G(3)$ using elementary row operations:

$$(3I - A)^2 = \begin{bmatrix} 2 & 1 & 0 \\ -2 & -1 & 0 \\ -2 & -1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

resulting in a basis:

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

As one might expect, this consists of two column vectors, because the dimension of a generalized eigenspace $G(\lambda)$ equals the algebraic multiplicity of λ (see proof of Corollary 1.5.4, which follows).¹⁷ Because $\lambda_2 = 4$ has multiplicity 1, the second generalized eigenspace $G(4)$ is simply the usual eigenspace $E(4)$. A simple calculation shows that this has

$$\left\{ \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \right\}$$

as a basis. The three displayed column vectors form a basis for F^3 , and confirm the generalized eigenspace decomposition $F^3 = G(3) \oplus G(4)$. On the other hand, since the eigenspace $E(3)$ of the eigenvalue $\lambda_1 = 3$ is only 1-dimensional, with basis

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix} \right\},$$

in contrast we have $F^3 \neq E(3) \oplus E(4)$ (equivalently, A is not diagonalizable). \square

Corollary 1.5.4 (Reduction to the Nilpotent Case)

Let $A \in M_n(F)$ where F is algebraically closed. Let $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of A . Then A is similar to a block diagonal matrix

$$B = \begin{bmatrix} \boxed{B_1} & & & \\ & \boxed{B_2} & & \\ & & \ddots & \\ & & & \boxed{B_k} \end{bmatrix} = \text{diag}(B_1, B_2, \dots, B_k)$$

such that each $B_i = \lambda_i I + N_i$ where N_i is a nilpotent matrix. Moreover, the size of the block B_i is the algebraic multiplicity of λ_i .

17. It's just as well this holds. It saves us some awkward terminology, such as "generalized geometric multiplicity of an eigenvalue."

Proof

Let \mathcal{B} be the standard basis for F^n and let T be the left multiplication map of A on column vectors. Let \mathcal{B}' be a basis consisting of a union of bases \mathcal{G}_i of the $G(\lambda_i)$. By Proposition 1.3.2 and the fact that each $G(\lambda_i)$ is T -invariant, A is similar to a block diagonal matrix $B = \text{diag}(B_1, B_2, \dots, B_k)$ where B_i is the matrix of T restricted to $G(\lambda_i)$ and relative to \mathcal{G}_i . In fact, $B = C^{-1}AC$ for $C = [\mathcal{B}', \mathcal{B}]$. Since $G(\lambda_i) = \ker(\lambda_i I - A)^{m_i}$, where m_i is the algebraic multiplicity of λ_i , we have that $(\lambda_i I - T)^{m_i}$ is zero on $G(\lambda_i)$. Hence, $(\lambda_i I - B_i)^{m_i} = 0$ and so $B_i = \lambda_i I + N_i$ where $N_i = B_i - \lambda_i I$ is nilpotent.

If the block B_i is an $h_i \times h_i$ matrix, then the characteristic polynomial of B is $(x - \lambda_1)^{h_1} (x - \lambda_2)^{h_2} \dots (x - \lambda_k)^{h_k}$. Inasmuch as the characteristic polynomial of A is the same as that of B , and has a unique irreducible factorization, h_i must be the algebraic multiplicity of λ_i . \square

Corollary 1.5.5

In the notation of the previous corollary and its proof, and using the same block structure as in B , if we form the block diagonal matrices

$$D = \left[\begin{array}{c|c|c} \lambda_1 I & & \\ \hline & \lambda_2 I & \\ & & \ddots \\ & & & \lambda_k I \end{array} \right], \quad N = \left[\begin{array}{c|c|c} N_1 & & \\ \hline & N_2 & \\ & & \ddots \\ & & & N_k \end{array} \right]$$

then D is diagonal, N is nilpotent, and $B = D + N$. Also D and N commute. Pulling these matrices back under the inverse conjugation, we have $A = CBC^{-1} = CDC^{-1} + CNC^{-1}$, which expresses a general matrix A as a sum of a diagonalizable matrix and a nilpotent matrix, which commute.

Remark 1.5.6

While on the subject of idempotent and nilpotent matrices,¹⁸ we note that a good stock of basic examples often comes in handy, as when testing conjectures. (For

18. We don't wish to give the impression here that invertible matrices are not important, outside of their role in similarity transformations. Although every matrix (over any field) is the sum of an idempotent matrix and an invertible matrix, this doesn't help much in understanding the structure of a general matrix, because we don't know the structure of the invertibles (without going back to nilpotents). However, under multiplication, the invertible $n \times n$ matrices form the general linear group $GL_n(F)$. Its rich group structure, harnessed through group representation theory and character theory, as developed by Frobenius, Burnside, Schur, Brauer and others, provides the most powerful tool known for describing finite groups and other classes of groups. There is also a sort of complementary multiplicative relationship of invertible and idempotent matrices. A 1967 result of J. A. Erdos says that every noninvertible matrix is a product of idempotent matrices.

instance, if E is a nontrivial (i.e., nonzero, nonidentity) idempotent matrix and N is nilpotent, does $E + N$ have two distinct eigenvalues? If E and N commute, are 0 and 1 then the only eigenvalues of $E + N$?) Of course, the canonical forms describe all idempotent and nilpotent matrices to within similarity, in a simple and beautiful way. But real-life examples won't always appear this way. (And, besides, a canonical form will not usually present the nice picture of two matrices simultaneously.) For 2×2 matrices, nontrivial idempotent matrices are precisely those with rank 1 and trace 1. And the nontrivial (i.e., nonzero) nilpotent matrices are those of rank 1 and trace 0. (In each case, the minimal polynomial, respectively $x^2 - x$ and x^2 , will agree with the characteristic polynomial here, which for a 2×2 matrix A is $x^2 - (\text{tr } A)x + \det A$.) For instance,

$$\begin{bmatrix} -2 & -1 \\ 6 & 3 \end{bmatrix}, \begin{bmatrix} -3 & -1 \\ 9 & 3 \end{bmatrix}$$

are respectively idempotent and nilpotent. □

1.6 SYLVESTER'S THEOREM ON THE MATRIX EQUATION

$$AX - XB = C$$

On several occasions throughout the book (beginning with our next proposition), we will invoke Sylvester's theorem¹⁹ on solutions to a matrix equation $AX - XB = C$. Operator theorists often refer to the result as Rosenblum's theorem.²⁰

Theorem 1.6.1 (Sylvester's Theorem)

Let F be an algebraically closed field. Let A and B be $n \times n$ and $m \times m$ matrices over F , respectively. If A and B have no eigenvalues in common, then for each $n \times m$ matrix C , the equation

$$(*) \quad AX - XB = C$$

has a unique solution $X \in M_{n \times m}(F)$.

Proof

We follow the 1959 proof of Lumer and Rosenblum as elegantly presented in the 1997 paper of Bhatia and Rosenthal.

19. Sylvester discovered the result in 1884. The theorem is well-known but deserves to be even better known, if for no other reason than its proof highlights the power of switching back and forth between matrices and linear transformations.

20. The operator version was first noted in the late 1940's and independently published by Dalecki in 1953 and Rosenblum in 1956.

Let $V = M_{n \times m}(F)$ and regard V as an mn -dimensional vector space over F . Let $T_A : V \rightarrow V$ and $T_B : V \rightarrow V$ be the left and right multiplication maps by A and B , respectively:

$$T_A(X) = AX, \quad T_B(X) = XB \quad \text{for all } X \in V.$$

Then T_A and T_B are commuting linear transformations of V (from associativity of matrix multiplication). As such, T_A and T_B can be simultaneously triangularized, that is, there is a basis in which both their matrices are upper triangular. (See Proposition 2.3.4 for a proof of this well-known fact.) In particular, since the eigenvalues of a triangular matrix are its diagonal entries, the eigenvalues of $T_A - T_B$ are differences of eigenvalues of T_A and eigenvalues of T_B . But the eigenvalues of T_A are the same as the eigenvalues of A (consider the columns of $X \in M_{n \times m}(F)$ in an equation $AX = \lambda X$), and the eigenvalues of T_B are the eigenvalues of B . Inasmuch as A and B have no common eigenvalues, the eigenvalues of $T_A - T_B$ are therefore all nonzero. Therefore, $T_A - T_B$ is an invertible transformation. The rest is easy:

Let $T = T_A - T_B$. The solutions to $(*)$ are exactly the solutions to

$$T(X) = C.$$

Invertibility of T implies there is a unique solution $X = T^{-1}(C)$. □

We record the following proposition for future use in connection with establishing uniqueness of the Weyr (or Jordan) canonical form. It says that the blocks B_i in Corollary 1.5.4 are unique to within similarity.

Proposition 1.6.2

Let $A = \text{diag}(A_1, A_2, \dots, A_k)$ and $B = \text{diag}(B_1, B_2, \dots, B_k)$ be similar block diagonal matrices over an algebraically closed field such that A_i and B_i have the same single eigenvalue λ_i , but $\lambda_i \neq \lambda_j$ when $i \neq j$. Then the matrices A_i and B_i must be of the same size and similar for $i = 1, 2, \dots, k$.

Proof

Inasmuch as A and B are similar, they must have the same characteristic polynomial, say $p(x) = (x - \lambda_1)^{m_1} (x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}$. Moreover, from our eigenvalue hypotheses, A_i and B_i are then $m_i \times m_i$ matrices. Let P be an invertible matrix with $P^{-1}AP = B$. As a $k \times k$ blocked matrix (with diagonal blocks matching those of A and B), write $P = (P_{ij})$.

We show that the off-diagonal blocks of P are zero. Fix indices i, j with $i \neq j$. From $P^{-1}AP = B$ we have $AP = PB$, whence

$$A_i P_{ij} = P_{ij} B_j.$$

But from Sylvester's Theorem 1.6.1 (taking $C = 0$), this implies $P_{ij} = 0$ because A_i and B_j have no common eigenvalue. Thus, $P = \text{diag}(P_{11}, P_{22}, \dots, P_{kk})$ is block diagonal.

From the way in which block diagonal matrices multiply, we now see that each P_{ii} is invertible and $P_{ii}^{-1}A_iP_{ii} = B_i$. Thus, A_i and B_i are similar, as our proposition claims. \square

1.7 CANONICAL FORMS FOR MATRICES

The theme of this book is a particular canonical form, the Weyr form, for square matrices over an algebraically closed field. It is a canonical form with respect to the equivalence relation of similarity. The rational form and Jordan form are also canonical forms for the same equivalence relation. How can they all be right? For that matter, what is a canonical form?

Let us illustrate the concept with the class of all $m \times n$ matrices, for fixed m and n , and over some fixed but arbitrary field, and with respect to the equivalence relation \sim of row equivalence: for $m \times n$ matrices A and B , $A \sim B$ if B can be obtained from A by elementary row operations. This is the same thing as A and B having the same row space. In this setting, the undisputed king of canonical forms is the reduced row-echelon form R of a matrix A :

$$R = \left[\begin{array}{cccccccc} 1 & * & 0 & 0 & * & * & 0 & * \\ & & 1 & 0 & * & * & 0 & * \\ & & & 1 & * & * & 0 & * \\ & & & & & & 1 & * \end{array} \right].$$

Note some properties of R :

- (1) For a given matrix A , there is a unique R in reduced row-echelon form such that $A \sim R$.
- (2) This unique R can be computed by an algorithm.
- (3) R “looks nice,” in this case with a lot of 0's, and 1's in a “staircase”²¹ formation.
- (4) Questions concerning the row space of R (and therefore of A) can be immediately answered. (For instance, the nonzero rows of R form a basis for the row space.)

21. Hence the term “echelon,” which describes a type of a military troop formation.

Of course, one could come up with other contrived canonical forms, such as requiring the leading entries of the nonzero rows of R to be 9's, say, instead of 1's (over a field of characteristic zero). So the reduced row-echelon form is not the only candidate for a canonical form, but it is surely the nicest.²²

There are also many things that R does not tell us about A . It won't tell us its determinant, trace, or eigenvalues if A is, say, a square matrix. But why should it? – these things are not invariants under the particular equivalence relation we are considering.

Now suppose we fix n and choose the similarity relation \sim on the class of $n \times n$ matrices over some fixed field F . One might then expect that a good canonical form for this equivalence relation \sim would satisfy:

- (1) Each $A \in M_n(F)$ is similar to a unique matrix in canonical form.
- (2) The canonical form of A , together with an explicit similarity transformation, can be computed by an algorithm.
- (3) The canonical form “looks nice.”
- (4) Computations with the canonical form, such as evaluating a polynomial expression, are relatively simple.
- (5) Questions about any standard invariant relative to similarity can be immediately answered for the canonical form (and therefore for the matrix A). For instance, the determinant, characteristic and minimal polynomials, eigenvalues and eigenvectors, should ideally be immediately recoverable from the form.

In short, a canonical form with respect to similarity should provide an exemplar for each similarity class of matrices—one particularly pleasant landmark for each similarity class, if you will. One should be able to more simply answer questions about a general matrix A by going to its canonical form via a similarity transformation, answering the question for the canonical matrix, and returning with an answer for A via the inverse similarity transformation.

The reason why there are several canonical forms in the market for the similarity relation is that they each meet the five stated goals in varying degrees, but without a clear overall winner. The three principal players are the rational, Jordan, and Weyr forms. Of the three, Jordan is the best known and Weyr the least. We won't give an account of the rational form. Many standard texts do.

22. Just when we think we know all about matrices in reduced row-echelon form, something new comes along, like this: The product of two $n \times n$ matrices in reduced row-echelon form is again in reduced row-echelon form. This surprising little result was recently pointed out to us by Vic Camillo, who used the result in his 1997 paper. However, Vic does not expect to have been the first to observe this and has asked for an earlier reference, perhaps an exercise in some linear algebra text.

We shall briefly describe the Jordan form later in this section. Our account of the Weyr form will begin in Chapter 2.

The Jordan and Weyr forms require an algebraically closed field, which dents objective (1) a little, but one can always pass to the algebraic closure of the base field. All three forms meet the uniqueness requirement, modulo a trivial variation. However, the Jordan and Weyr forms meet (2) only in theory. Each requires that the eigenvalues of the matrix A be known. When this is the case, then the Weyr form has a simpler algorithm than the Jordan form, not only in terms of calculating what the canonical form of the original matrix A will be, but also in computing a similarity transformation. On the other hand, the rational form really can be computed algorithmically over any field (hence its name). That is its great strength. In the beauty stakes (3), the authors would judge Jordan the winner, Weyr second, and rational third.²³ The Jordan form also works very well in (4) when working in *isolation*. But we will see that, in its *interactions* with other matrices, the Weyr form clearly has the upper hand. For (5), Jordan and Weyr tie for first, with the rational a distant third (apart from the minimal and characteristic polynomials, not much else is evident from the rational form, most notably no eigenvalues).

It is a mistake, however, to view the three forms as being in “competition” with each other. All are worthy. A particular form may be better in some circumstances, and inferior in others. One should be prepared to switch back and forth, according to the situation.

The 1932 text *An Introduction to the Theory of Canonical Matrices* by Turnbull and Aitken²⁴ is still a classic, but perhaps a little hard for the modern reader to appreciate, because of its outdated mathematical language and terminology. The text does mention Eduard Weyr’s work, specifically the *Weyr characteristic* (which we later term “Weyr structure”). It is one of the few books occasionally referenced as a source for the Weyr canonical form. This is mistaken—the Weyr canonical form itself is not (explicitly) covered in the book. Undoubtedly, the authors knew of the canonical form. It says a lot that, in the space of some 45 years after Weyr’s discovery of his form in 1885, the form had been largely dismissed—not mentioned in even a linear algebra book devoted to canonical forms. However, applications of the form

23. Of course, these are subjective judgments.

24. A brilliant mathematician, Aitken received his undergraduate education at the University of Otago (New Zealand), and later became Professor of Mathematics at the University of Edinburgh. The second author of our book is a faculty member at the University of Otago. The first author began his undergraduate linear algebra education at Otago under fellow New Zealander J. S. (James) Milne. A number theorist, Milne has contributed at the highest level, based mostly at the University of Michigan, Ann Arbor. In Chapter 7, we reference his notes in algebraic geometry.

itself, for instance to commutativity problems in matrix theory.²⁵ may not have been appreciated then.

A good canonical form under similarity allows an indirect way of seeing if two given matrices $A, B \in M_n(F)$ are similar, by checking if they have the same canonical form. This is sometimes highlighted by authors as the *raison d'être* for having a canonical form. That the present authors chose not to list this outcome among their five desirable features of a canonical form would suggest they disagree! Nevertheless, the Jordan and Weyr forms both perform well in this method of testing for similarity of matrices A and B (with perhaps Weyr a little more transparent), providing one knows the eigenvalues of A and B (which must agree for similar matrices), and also knows the nullities of the various powers $(\lambda I - A)^i$ and $(\lambda I - B)^i$ for $i = 1, 2, \dots, n$ as λ ranges over the eigenvalues. One then simply checks that these nullities agree. (The details are given in Chapter 2, Proposition 2.2.8.)

We next briefly outline the Jordan form of a square matrix A over an algebraically closed field F , such as the complex numbers \mathbb{C} . The material is covered in many books. See, for instance, the linear algebra texts by Hoffman and Kunze, Horn and Johnson, Nicholson, and Weintraub, listed in our bibliography. A **basic Jordan matrix with eigenvalue** λ takes the form

$$J = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & \ddots \\ & & & & \lambda & 1 \\ & & & & & \lambda \end{bmatrix}.$$

It has the eigenvalue λ repeated down the main diagonal, has 1's on the first superdiagonal, and all other entries are 0. Note that $J = \lambda I + N$, where N is the basic nilpotent Jordan matrix

$$N = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & \ddots \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix}.$$

25. This is now an actively researched area, as evidenced in our bibliography.

In terms of linear transformations, N is the matrix of the quintessential nilpotent linear transformation $T : V \rightarrow V$ of an n -dimensional space V whose action relative to some basis $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ is the backward shifting

$$0 \longleftarrow v_1 \longleftarrow v_2 \longleftarrow v_3 \longleftarrow \cdots \longleftarrow v_{n-1} \longleftarrow v_n.$$

This transformation has index of nilpotency n and its matrix $[T]_{\mathcal{B}}$ is precisely N . Thus, a basic Jordan matrix is a scalar matrix plus the nicest possible nilpotent matrix. Note in particular how the powers of N behave.

A **Jordan matrix** J in general is simply a direct sum of basic Jordan matrices, without restrictions on the associated eigenvalues, the number of basic blocks associated with a given eigenvalue, or their sizes:

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{bmatrix} = \text{diag}(J_1, J_2, \dots, J_k),$$

where each J_i is a basic Jordan matrix for some associated eigenvalue λ_i . For a common eigenvalue λ , we agree to group together its corresponding blocks, and in decreasing order of size. If these basic block sizes are $m_1 \geq m_2 \geq \dots \geq m_s$, we call (m_1, m_2, \dots, m_s) the **Jordan structure** of J for the associated eigenvalue λ . (This is also known as the **Segre characteristic**.) Every square matrix A over F is similar to a Jordan matrix J , which is unique to within the ordering of the blocks determined by our chosen ordering of the eigenvalues. We call J the **Jordan canonical form**, or simply the **Jordan form**, of A . The **Jordan structure** of A associated with an eigenvalue λ is then defined as the corresponding Jordan structure of J associated with λ . It is clear from Corollary 1.5.4, that it is enough to establish the Jordan form for an $n \times n$ nilpotent matrix. We refer the reader to a standard text for the details. However, the existence and uniqueness of the Jordan form also follow from our work on the Weyr form in Chapter 2.

Example 1.7.1

The following matrix J is in Jordan form, with two distinct eigenvalues 2 and 6, and corresponding Jordan structures $(3, 3, 2, 1, 1)$ and $(2, 2)$,

respectively:

$$J = \left[\begin{array}{ccc|ccc|cc|c|cc} 2 & 1 & 0 & & & & & & & & \\ 0 & 2 & 1 & & & & & & & & \\ 0 & 0 & 2 & & & & & & & & \\ \hline & & & 2 & 1 & 0 & & & & & \\ & & & 0 & 2 & 1 & & & & & \\ & & & 0 & 0 & 2 & & & & & \\ \hline & & & & & & 2 & 1 & & & \\ & & & & & & 0 & 2 & & & \\ & & & & & & & & 2 & & \\ & & & & & & & & & 2 & \\ & & & & & & & & & & 6 & 1 \\ & & & & & & & & & & 0 & 6 \\ & & & & & & & & & & & 6 & 1 \\ & & & & & & & & & & & 0 & 6 \end{array} \right] \quad \square$$

Some authors principally motivate the Jordan form by saying that it is an “almost diagonal” matrix. That property is certainly nice aesthetically, and the sparseness of the nonzero entries therein is computationally desirable.²⁶ But this view seems to miss a much more important point connected with the fourth desirable feature we listed for a canonical form. Namely, the beautiful shifting effect²⁷ that a basic nilpotent Jordan matrix J has when it right (resp. left) multiplies any other matrix A , in terms of what happens to the columns (resp. rows) of A in the product AJ (resp. JA). (The reader may wish to try and pick this pattern using a small matrix A .) This makes for easy computations. In particular, it is a breeze to compute a polynomial $c_0I + c_1J + c_2J^2 + \cdots + c_kJ^k$ in a basic nilpotent Jordan matrix. For instance, when J is 5×5 , we have:

$$c_0I + c_1J + c_2J^2 + c_3J^3 + c_4J^4 = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 & c_4 \\ & c_0 & c_1 & c_2 & c_3 \\ & & c_0 & c_1 & c_2 \\ & & & c_0 & c_1 \\ & & & & c_0 \end{bmatrix}.$$

26. The Jordan form is actually optimal with respect to the number of nonzero off-diagonal entries (the Weyr form also shares this property), although not necessarily optimal in the total number of nonzero entries of similar matrices. See the 2008 article by Brualdi, Pei, and Zhan.

27. The authors suspect Jordan was more motivated by this shifting phenomenon, perhaps in terms of transformations, than the desire to get an almost diagonal matrix.

The shifting behavior of both the Jordan and Weyr forms is discussed in depth in Chapter 2. In general, as we will see, the Weyr form is certainly not “almost diagonal,” until it is viewed as a blocked matrix. It incorporates shifting more universally than its Jordan cousin, *but in terms of blocked matrices*.

In view of uniqueness of the Jordan form, there are as many dissimilar $n \times n$ nilpotent matrices as there are possible Jordan structures, that is, the number $\pi(n)$ of partitions²⁸ of n . (And therefore exactly $\sum \pi(n_1)\pi(n_2) \cdots \pi(n_k)$ dissimilar $n \times n$ matrices having k given distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$, where the summation is over all ordered k -tuples (n_1, n_2, \dots, n_k) of positive integers whose entries sum to n .)²⁹ For instance, up to similarity there are exactly $\pi(5) = 7$ nilpotent 5×5 matrices. Here they are:

Example 1.7.2

The seven 5×5 nilpotent Jordan matrices.

$$\left[\begin{array}{c|c|c|c|c} 0 & & & & \\ \hline & 0 & & & \\ & & 0 & & \\ & & & 0 & \\ & & & & 0 \end{array} \right] \quad \text{Jordan structure } (1, 1, 1, 1, 1)$$

$$\left[\begin{array}{cc|c|c|c} 0 & 1 & & & \\ 0 & 0 & & & \\ \hline & & 0 & & \\ & & & 0 & \\ & & & & 0 \end{array} \right] \quad \text{Jordan structure } (2, 1, 1, 1)$$

$$\left[\begin{array}{cc|cc|c} 0 & 1 & & & \\ 0 & 0 & & & \\ \hline & & 0 & 1 & \\ & & 0 & 0 & \\ & & & & 0 \end{array} \right] \quad \text{Jordan structure } (2, 2, 1)$$

28. By a **partition of n** we nearly always mean a finite sequence (n_1, n_2, \dots, n_r) of positive integers such that $n = n_1 + n_2 + \cdots + n_r$ and $n_1 \geq n_2 \geq \cdots \geq n_r$, that is, an ordered partition of n with decreasing parts. This always applies to the Jordan and Weyr structures of matrices. However, with minimal confusion, in other settings we very occasionally dispense with the decreasing requirement, as for instance in Proposition 1.2.1 and the discussion preceding it. The context should always make this clear.

29. Try getting this answer without canonical forms! Powerful tools.

$$\left[\begin{array}{ccc|c} 0 & 1 & 0 & \\ 0 & 0 & 1 & \\ 0 & 0 & 0 & \\ \hline & & & 0 \\ & & & \hline & & & 0 \end{array} \right]$$

Jordan structure (3, 1, 1)

$$\left[\begin{array}{ccc|c} 0 & 1 & 0 & \\ 0 & 0 & 1 & \\ 0 & 0 & 0 & \\ \hline & & & 0 & 1 \\ & & & 0 & 0 \end{array} \right]$$

Jordan structure (3, 2)

$$\left[\begin{array}{cccc|c} 0 & 1 & 0 & 0 & \\ 0 & 0 & 1 & 0 & \\ 0 & 0 & 0 & 1 & \\ 0 & 0 & 0 & 0 & \\ \hline & & & & 0 \end{array} \right]$$

Jordan structure (4, 1)

$$\left[\begin{array}{ccccc} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

Jordan structure (5)

BIOGRAPHICAL NOTES ON JORDAN AND SYLVESTER

Camille Jordan was born on January 5, 1838, in Lyon, France, the son of an engineer. In 1855 he entered the École Polytechnique, Paris, to study mathematics. He also trained there as an engineer and this became his profession. However, he was clearly able to both work as an engineer and spend considerable time on mathematical research. He was examined for his doctorate in 1861 and his thesis came in two parts: *Sur le nombre des valeurs des fonctions*, on algebra, and *Sur des périodes des fonctions inverses des intégrales des différentielles algébriques*. As a mathematician, he worked in a wide range of areas making valuable contributions to every mathematical topic of his day. He was particularly interested in finite groups, although he viewed these as groups of permutations. His work between 1860 and 1870 in this area was published in 1870 in what can be thought of as the first group theory book, titled *Traité des Substitutions et des Équations Algébriques*. Book 2 of this treatise contains the Jordan form, although only for matrices over a finite field, not \mathbb{C} , and couched

in the language of permutation theory instead of matrices. He became professor of analysis at the École Polytechnique in 1876 and professor at the Collège de France in 1883 and was Honorary President of the International Congress of Mathematicians at Strasbourg in 1920. He died in Paris on January 22, 1922.

James Joseph Sylvester was born in London on September 3, 1814, the son of a merchant. He was brought up in the Jewish faith and this was an impeding factor in his academic life. While just 14, he entered the nonsectarian University College, London, in 1828. However five months later his family withdrew him from the college after he was accused of threatening a fellow student with a knife. In 1831 he began studies at St John's College, Cambridge, and, after a period of illness, sat the mathematical tripos (final) examination in 1837, coming second in his class. However, graduation required swearing an oath to the Church of England and Sylvester's religion prevented this. The University of London had no religious hurdles and, although he had no degree, Sylvester was awarded the chair of natural philosophy there in 1838. Then, just 27 years old, he was appointed to the chair of mathematics at the University of Virginia in the United States. However, he resigned from this position after only a few months following an incident in which he struck a student in his class with a sword stick. He returned to London but found it difficult to get an academic position. Instead he became a lawyer. Fortunately, Arthur Cayley was also practicing law at this time and they became good friends, meeting at the law courts to discuss mathematics. In the mid-1850s Sylvester eventually secured a professorship at the Royal Military Academy at Woolwich, London. He did extensive work in matrix theory, indeed was responsible for the term "matrix" (as well as "derogatory") and used matrix theory in higher dimensional geometry. Although he retired from Woolwich in 1870, in 1877 he accepted a chair at Johns Hopkins University in the United States and in 1878 he founded the *American Journal of Mathematics*, the first U.S. mathematics journal. Then, aged 68, he was appointed to the Savilian Chair of Geometry at Oxford, a position he kept for 10 years. He died in London on March 15, 1897.

The Weyr Form

Here enters the principal actor. Our aim in this chapter is to describe the Weyr form and its basic properties to a reader we shall assume has never heard of the Weyr form but is moderately familiar with the Jordan canonical form. We delay applications of the Weyr form until Part II of the book (Chapters 5, 6, and 7).

Very few people, even specialists in linear algebra, know of the Weyr form. The Czech mathematician Eduard Weyr discovered the form in 1885. In the intervening 125 years, the form has been rediscovered periodically, under various names (such as “modified Jordan form,” “reordered Jordan form,” “second Jordan form,” and “H-form”). No doubt those authors attempted to convey their enthusiasm for the form to others, but the Weyr form has never really caught on. Possibly several factors have been at play here. First, the expository accounts of the Weyr form have often left a lot to be desired. Second, it is very easy to miss the point of the Weyr form, even if one knows what the form is. In matrix terms, the Jordan form looks nicer. So why change? Third, some have mistakenly interpreted a duality between the Jordan and Weyr forms as saying the Weyr form is a “mere” permutation of the Jordan form, whereas in fact there is a big conceptual difference in the two forms. And fourth, some folk, perhaps through poor motivation or not feeling entirely comfortable with

blocked matrices, fail to fully grasp the concept. Yet when one “gets it,” the Weyr form is so simple, natural, and useful. It can be fully understood by anyone who has the background to understand the Jordan form. In many applications, the Weyr form is superior to the Jordan form. Moreover, as we show in Chapter 4, the Weyr form, as a mathematical concept, lives in a somewhat bigger universe than its Jordan counterpart.

Given this background, the authors were conscious of the need for great care with the description of the Weyr form in this chapter. If anything, perhaps we have tended to err on the side of “over-measure.”¹ The path to understanding the Weyr form can take different routes, depending on the background of the individual reader. Our advice to the reader is to proceed at his or her own pace. Skip details if they seem “obvious” and re-read if they are not. For instance, we would recommend to readers who do not feel entirely comfortable with the Weyr form by chapter’s end, that they re-work the calculations (preferably by hand) for computing the Weyr form of specific matrices, given in Section 2.5. (And do answer both test questions.) On the other hand, a more confident reader may prefer to simply skip or scan the examples and to return later if the need arises.

We now briefly outline the chapter’s contents. Section 2.1 presents some motivation for the Weyr form definitions, the definitions themselves, and numerous examples of matrices in Weyr form, but without developing properties of the Weyr form. Section 2.2 establishes that every square matrix over an algebraically closed field F is similar to a unique matrix in Weyr form. In Section 2.3, we show that given a list A_1, A_2, \dots, A_k of commuting $n \times n$ matrices over F , it is possible under a similarity transformation to put A_1 in Weyr form and simultaneously have A_2, A_3, \dots, A_k in upper triangular form. The first and third authors were led to the rediscovery of the Weyr form because of the desire for this property, which is not shared by the Jordan form. This extended triangularization result is a useful computational tool, which we use in Chapter 6 to study the question of when A_1, A_2, \dots, A_k can be approximated by simultaneously diagonalizable matrices. (The latter question has arisen in recent problems in biomathematics and multivariate interpolation.) Section 2.4 develops a nice duality between the Jordan forms and the Weyr forms of nilpotent matrices. Finally, Section 2.5 provides an algorithm for computing the Weyr form of a square matrix, along with instructive examples. We have made little attempt to address the question of the numerical stability of the algorithm. That question is important because one application of the Weyr

1. Although experience tells us that it is hard to overdo the Weyr form explanations with a fresh audience.

form (not covered in this book) is to the numerical linear algebra problem of computing the Jordan form in a stable manner.

2.1 WHAT IS THE WEYR FORM?

The goal of this section is to make readers feel comfortable with what the Weyr form is, but without attempting at this stage to say what its properties are or what it is good for. The bold among us could go straight to the formal definitions of the Weyr form given in Definitions 2.1.1 and 2.1.5 and proceed from there. The definitions are quite precise but yet one can easily miss the underlying concept. So let's begin with some motivation.

Fix an algebraically closed field F , and perhaps keep in mind the particular case of the field \mathbb{C} of complex numbers. As we will see, the Weyr form, like the Jordan form, for square matrices over F quickly comes down to the nilpotent case. By way of motivation, consider the nilpotent 10×10 matrix J in Jordan form and with Jordan structure $(4, 4, 2)$:

$$J = \left[\begin{array}{cccc|cccc|cc} 0 & 1 & 0 & 0 & & & & & & \\ & 0 & 1 & 0 & & & & & & \\ & & 0 & 1 & & & & & & \\ & & & 0 & & & & & & \\ \hline & & & & 0 & 1 & 0 & 0 & & \\ & & & & & 0 & 1 & 0 & & \\ & & & & & & 0 & 1 & & \\ & & & & & & & 0 & & \\ \hline & & & & & & & & 0 & 1 \\ & & & & & & & & & 0 \end{array} \right]$$

We can view J as the matrix of a linear transformation T of $V = F^{10}$ relative to some (ordered) basis $\mathcal{B} = \{v_1, v_2, \dots, v_{10}\}$ (e.g., the standard basis, in which case T acts under left multiplication by J on column vectors). The action of T on the basis vectors can be naturally represented as:

$$\begin{array}{ccccccccc} 0 & \leftarrow & v_1 & \leftarrow & v_2 & \leftarrow & v_3 & \leftarrow & v_4 \\ 0 & \leftarrow & v_5 & \leftarrow & v_6 & \leftarrow & v_7 & \leftarrow & v_8 \\ 0 & \leftarrow & v_9 & \leftarrow & v_{10} & & & & \end{array}$$

Here our attention is drawn to the three rows of the diagram, which correspond to the three Jordan blocks, and the cyclic² shifting and annihilation within each block. But now cast one's eyes to the columns. What are they revealing about T ? The first column after the zeros reveals that $\{v_1, v_5, v_9\}$ is a basis for the null space of T . The next column then reveals that $\{v_1, v_5, v_9, v_2, v_6, v_{10}\}$ is a basis for the null space of T^2 . And so on—the next two columns supplement the earlier ones to give bases for the null spaces of T^3 and T^4 , respectively. That could be useful, although it hasn't captured the way in which the supplements are mapped to the previous null space.

What happens if we reorder the basis vectors in \mathcal{B} by running through them in column order? We get the basis $\mathcal{B}' = \{v_1, v_5, v_9, v_2, v_6, v_{10}, v_3, v_7, v_4, v_8\} = \{v'_1, v'_2, \dots, v'_{10}\}$. Just as it was natural to view \mathcal{B} as a union of the three groupings corresponding to the rows (after dropping each 0), so it is natural to view \mathcal{B}' as a bunch of four groupings corresponding to the columns. In line with these groupings, but now written in four rows to conform to our initial Jordan view, the action of T on the primed basis vectors is:

$$\begin{array}{ccc}
 0 & 0 & 0 \\
 \uparrow & \uparrow & \uparrow \\
 v'_1 & v'_2 & v'_3 \\
 \uparrow & \uparrow & \uparrow \\
 v'_4 & v'_5 & v'_6 \\
 \uparrow & \uparrow & \\
 v'_7 & v'_8 & \\
 \uparrow & \uparrow & \\
 v'_9 & v'_{10} &
 \end{array}$$

This diagram, of course, is just the transpose of the Jordan diagram above. Note with the Jordan view, it was the “within-row” action which was interesting—there was no interaction between rows. With the new basis, it is the “interaction between rows” which strikes one. It is almost as if the various cyclic shiftings of the Jordan matrix on basis vectors within each of its three basic blocks have been replaced by a **single cyclic shift on the four subspaces spanned by the rows**. Grasp that, and you have grasped an important aspect of the Weyr

2. Perhaps “backwards shifting” or “left shifting” would be a more accurate term here, because the “cycle” is not completed (e.g., v_1 gets bumped off, not cycled back to v_4). However, “cyclic” shifting has a better “ring” to it.

form concept. The matrix of T in the new basis \mathcal{B}' is:

$$W = \left[\begin{array}{ccc|ccc|cc|cc} 0 & 0 & 0 & 1 & 0 & 0 & & & & \\ 0 & 0 & 0 & 0 & 1 & 0 & & & & \\ 0 & 0 & 0 & 0 & 0 & 1 & & & & \\ \hline & & & 0 & 0 & 0 & 1 & 0 & & \\ & & & 0 & 0 & 0 & 0 & 1 & & \\ & & & 0 & 0 & 0 & 0 & 0 & & \\ \hline & & & & & & 0 & 0 & 1 & 0 \\ & & & & & & 0 & 0 & 0 & 1 \\ \hline & & & & & & & & 0 & 0 \\ & & & & & & & & 0 & 0 \end{array} \right]$$

This matrix W will turn out to be the Weyr form of J . Since W and J are matrices of the same transformation T under reordering of a basis, we must have $W = P^{-1}JP$ where P is the corresponding permutation matrix.

The linear transformation point of view for the Weyr form will be studied in considerable depth and generality in Chapter 4 (for those readers who are interested—it is optional). For the most part, however, it is the straight matrix view of the Weyr form which is of principal interest to us (and which is relevant to our applications).³ Of course, that is not to say that one should ignore the transformation viewpoint.⁴

Continuing our informal discussion, we next motivate the Weyr form of a matrix having a single eigenvalue (ignoring multiplicities) as a natural blocked matrix analogue of a *basic* Jordan matrix. If we take, say, the 6×6 nilpotent Jordan matrix of Jordan structure $(3, 3)$

$$J = \left[\begin{array}{ccc|ccc} 0 & 1 & 0 & & & \\ 0 & 0 & 1 & & & \\ 0 & 0 & 0 & & & \\ \hline & & & 0 & 1 & 0 \\ & & & 0 & 0 & 1 \\ & & & 0 & 0 & 0 \end{array} \right],$$

3. Also most people primarily think of the Jordan form in terms of matrices.

4. Any mathematician worth his or her salt continually flips back and forth between matrices and transformations in linear algebra problems.

and rework our earlier calculations of reordering a basis, we find, as the reader should quickly verify, that the Weyr form of J is

$$W = \left[\begin{array}{cc|cc|cc} 0 & 0 & 1 & 0 & & \\ 0 & 0 & 0 & 1 & & \\ \hline & & 0 & 0 & 1 & 0 \\ & & 0 & 0 & 0 & 1 \\ \hline & & & & 0 & 0 \\ & & & & 0 & 0 \end{array} \right].$$

Written as a blocked matrix with 2×2 blocks,

$$W = \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & I \\ 0 & 0 & 0 \end{bmatrix}.$$

This looks exactly like a 3×3 *basic* nilpotent Jordan matrix, except the 0's and 1's have been replaced by their 2×2 equivalents. It also suggests defining a **basic Weyr matrix** as a blocked matrix generalization of a *basic* Jordan matrix

$$J = \begin{bmatrix} \lambda & 1 & & & & \\ & \lambda & 1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & \lambda & 1 \\ & & & & & \lambda \end{bmatrix}.$$

In the blocked (Weyr) form, we replace the λ 's by scalar matrices λI

$$\lambda I = \begin{bmatrix} \lambda & & & & \\ & \lambda & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & \lambda \end{bmatrix},$$

and we replace the 1's by various "identity matrices" I . If we were to insist that all the identity matrices involved have the same size, that would cover our 6×6 example of a Weyr matrix W but not our original 10×10 example. However, observe that the 10×10 Weyr matrix W has the following form: its diagonal

blocks are of the form λI , for $\lambda = 0$ (the sole eigenvalue of W), and are of decreasing order of size 3×3 , 3×3 , 2×2 , 2×2 going down the diagonal blocks. And the first superdiagonal blocks of W are of the form

$$\begin{bmatrix} I \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & & \\ & 0 & 0 & \dots & 1 \\ \vdots & & & & 0 \\ & & & & \vdots \end{bmatrix},$$

that is, an identity matrix followed by zero rows (if the block is not square). Such matrices are characterized as being in reduced row-echelon form and having full column-rank. (Notice that the sizes of the off-diagonal blocks are necessarily dictated by the diagonal block sizes.) All other superdiagonal blocks of W are zero. It is time to formalize this definition of a basic Weyr matrix, or equivalently, the definition of a Weyr matrix having only a single eigenvalue. This will be followed by numerous examples, and then the easy definition of a general Weyr matrix.

Definition 2.1.1: A **basic Weyr matrix with eigenvalue** λ is an $n \times n$ matrix W of the following form: There is a partition $n_1 + n_2 + \dots + n_r = n$ of n with $n_1 \geq n_2 \geq \dots \geq n_r \geq 1$ such that, when W is viewed as an $r \times r$ blocked matrix (W_{ij}) , where the (i, j) block W_{ij} is an $n_i \times n_j$ matrix, the following three features are present:

- (1) The main diagonal blocks W_{ii} are the $n_i \times n_i$ scalar matrices λI for $i = 1, \dots, r$.
- (2) The first superdiagonal blocks $W_{i, i+1}$ are full column-rank $n_i \times n_{i+1}$ matrices in reduced row-echelon form (that is, an identity matrix followed by zero rows) for $i = 1, \dots, r - 1$.
- (3) All other blocks of W are zero (that is, $W_{ij} = 0$ when $j \neq i, i + 1$).

In this case, we say that W has **Weyr structure** (n_1, n_2, \dots, n_r) .

Example 2.1.2

The following seven matrices are all basic Weyr matrices, each with the eigenvalue λ . Next to each is recorded its Weyr structure. The readers are encouraged to discover for themselves a quick way of deducing the Jordan structures of the corresponding matrices in Jordan form. (This connection will be established

formally in Section 2.4.)

$$\left[\begin{array}{cccc|cc} \lambda & 0 & 0 & 0 & 1 & 0 \\ & \lambda & 0 & 0 & 0 & 1 \\ & & \lambda & 0 & 0 & 0 \\ & & & \lambda & 0 & 0 \\ \hline & & & & \lambda & 0 \\ & & & & & \lambda \\ & & & & & & 1 & 0 \\ & & & & & & & 1 \\ & & & & & & & & \lambda & 0 \\ & & & & & & & & & \lambda & 1 \\ & & & & & & & & & & \lambda \end{array} \right]$$

Weyr structure (4, 2, 2, 1, 1)

$$\left[\begin{array}{ccc|ccc} \lambda & 0 & 0 & 1 & 0 & 0 \\ & \lambda & 0 & 0 & 1 & 0 \\ & & \lambda & 0 & 0 & 1 \\ \hline & & & \lambda & 0 & 0 \\ & & & & \lambda & 0 \\ & & & & & \lambda \\ & & & & & & 1 & 0 \\ & & & & & & & 1 \\ & & & & & & & & \lambda & 0 \\ & & & & & & & & & \lambda & 1 \\ & & & & & & & & & & \lambda \end{array} \right]$$

Weyr structure (3, 3, 2, 2)

$$\left[\begin{array}{ccc|cc} \lambda & 0 & 0 & 1 & 0 \\ & \lambda & 0 & 0 & 1 \\ & & \lambda & 0 & 0 \\ \hline & & & \lambda & 0 \\ & & & & \lambda \\ & & & & & 1 & 0 \\ & & & & & & 1 \\ & & & & & & & \lambda & 0 \\ & & & & & & & & \lambda \end{array} \right]$$

Weyr structure (3, 2, 2)

$$\left[\begin{array}{ccc|ccc} \lambda & 0 & 0 & 1 & 0 & 0 \\ & \lambda & 0 & 0 & 1 & 0 \\ & & \lambda & 0 & 0 & 1 \\ \hline & & & \lambda & 0 & 0 \\ & & & & \lambda & 0 \\ & & & & & \lambda \end{array} \right]$$

Weyr structure (3, 3)

$$\left[\begin{array}{cc|cc|cc} \lambda & 0 & 1 & 0 & & \\ & \lambda & 0 & 1 & & \\ \hline & & \lambda & 0 & 1 & 0 \\ & & & \lambda & 0 & 1 \\ \hline & & & & \lambda & 0 \\ & & & & & \lambda \end{array} \right]$$

Weyr structure (2, 2, 2)

$$\left[\begin{array}{c|c|c|c} \lambda & 1 & & \\ \hline & \lambda & 1 & \\ & & \lambda & 1 \\ & & & \lambda \end{array} \right]$$

Weyr structure (1, 1, 1, 1)

$$\left[\begin{array}{cccc} \lambda & 0 & 0 & 0 \\ & \lambda & 0 & 0 \\ & & \lambda & 0 \\ & & & \lambda \end{array} \right]$$

Weyr structure (4)

□

Note that we can regard an $n \times n$ scalar matrix as a basic Weyr matrix with the trivial Weyr structure (n) . At the other extreme, a basic Jordan matrix is a basic Weyr matrix with Weyr structure $(1, 1, 1, \dots, 1)$. The Weyr structure (n_1, n_2, \dots, n_r) is called **homogeneous** if $n_1 = n_2 = \dots = n_r$. Basic Weyr $n \times n$ matrices with a homogeneous Weyr structure are the easiest to picture, because the “identity” blocks on the first superdiagonal are genuine $d \times d$ identity matrices for $d = n/r$.

Remark 2.1.3

In the case of the Jordan form, a Jordan matrix J with a single eigenvalue λ is a (direct) sum of basic Jordan matrices J_1, J_2, \dots, J_s , each having λ as its eigenvalue. In fact J_i is $m_i \times m_i$ where (m_1, m_2, \dots, m_s) is the Jordan structure of J . But in the case of the Weyr form, a Weyr matrix W with a single eigenvalue is NOT a (proper) sum of basic Weyr matrices. **By definition, a Weyr matrix with a single eigenvalue is the same thing as a basic Weyr matrix.** A direct sum of basic Weyr matrices with the same eigenvalue is not even a Weyr matrix according to the general definition. This is an important distinction between the Jordan and Weyr forms to bear in mind. We could have circumvented this possible confusion by not using the term “basic Weyr matrix” and instead using the more clumsy “a Weyr matrix with a single eigenvalue.” We have chosen not to, principally because we want to reinforce the idea that a basic Weyr matrix is a blocked matrix analogue of a basic Jordan matrix.

□

Before giving the definition of a Weyr matrix with more than one eigenvalue, we recall that (over an algebraically closed field) the nilpotent matrices are those having 0 as their only eigenvalue, and a (square) matrix A with a single eigenvalue λ is the same thing as $\lambda I + N$ where N is a nilpotent matrix (Proposition 1.1.1). So one should really view a Weyr matrix having a single eigenvalue as just a scalar matrix plus a nilpotent Weyr matrix. As with nilpotent Jordan matrices, there are exactly $\pi(n)$ nilpotent $n \times n$ Weyr matrices, where $\pi(n)$ is the number of partitions of n , because this is the number of possible Weyr structures. At the end of Chapter 1, we listed the $\pi(5) = 7$ nilpotent 5×5 Jordan matrices. Well, if it was good enough for Jordan, then it is good enough for Weyr. Here are the 5×5 nilpotent Weyr matrices, listed in the order that matches their Jordan counterparts to within similarity (e.g., the third Weyr matrix has the third Jordan matrix as its Jordan form):

Example 2.1.4

The seven 5×5 nilpotent Weyr matrices.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{Weyr structure (5)}$$

$$\left[\begin{array}{cccc|c} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline & & & & 0 \end{array} \right] \quad \text{Weyr structure (4, 1)}$$

$$\left[\begin{array}{ccc|cc} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ \hline & & & 0 & 0 \\ & & & 0 & 0 \end{array} \right] \quad \text{Weyr structure (3, 2)}$$

$$\left[\begin{array}{ccc|c|c} 0 & 0 & 0 & 1 & \\ 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & \\ \hline & & & 0 & 1 \\ & & & & 0 \end{array} \right] \quad \text{Weyr structure (3, 1, 1)}$$

$$\left[\begin{array}{cc|cc|c} 0 & 0 & 1 & 0 & \\ 0 & 0 & 0 & 1 & \\ \hline & & 0 & 0 & 1 \\ & & 0 & 0 & 0 \\ \hline & & & & 0 \end{array} \right]$$

Weyr structure (2, 2, 1)

$$\left[\begin{array}{cc|c|c|c} 0 & 0 & 1 & & \\ 0 & 0 & 0 & & \\ \hline & & 0 & 1 & \\ & & & 0 & 1 \\ & & & & 0 \end{array} \right]$$

Weyr structure (2, 1, 1, 1)

$$\left[\begin{array}{c|c|c|c|c} 0 & 1 & & & \\ \hline & 0 & 1 & & \\ & & 0 & 1 & \\ & & & 0 & 1 \\ & & & & 0 \end{array} \right]$$

Weyr structure (1, 1, 1, 1, 1)

□

With any reasonable matrix canonical form with respect to the similarity relation (and over an algebraically closed field), a matrix that is in canonical form and has $\lambda_1, \lambda_2, \dots, \lambda_k$ as its distinct eigenvalues must be a direct sum of k canonical matrices in which the i th summand has λ_i as its single eigenvalue. The Weyr form is no exception. Here is our definition of a general matrix in Weyr form.

Definition 2.1.5: Let W be a square matrix over an algebraically closed field F , and let $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of W . We say that W is in **Weyr form** (or is a **Weyr matrix**) if W is a direct sum of basic Weyr matrices, one for each distinct eigenvalue. In other words, W has the form

$$W = \left[\begin{array}{c|c|c|c|c} W_1 & & & & \\ \hline & W_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & W_k \end{array} \right]$$

where W_i is a basic Weyr matrix with eigenvalue λ_i for $i = 1, \dots, k$.

□

We again stress that a general Weyr matrix cannot have multiple basic Weyr blocks for the *same* eigenvalue.

Example 2.1.6

The following matrix W is in Weyr form, with two distinct eigenvalues 4 and 7, and corresponding Weyr structures $(2, 2, 1, 1, 1)$ and $(3, 2, 2)$ for its two basic blocks:

$$W = \left[\begin{array}{cccc|cccc|cccc|cccc|cccc} 4 & 0 & & & 1 & 0 & & & & & & & & & & & & \\ & 4 & & & 0 & 1 & & & & & & & & & & & & \\ & & & & 4 & 0 & 1 & & & & & & & & & & & \\ & & & & & 4 & 0 & & & & & & & & & & & \\ & & & & & & 4 & 1 & & & & & & & & & & \\ & & & & & & & 4 & 1 & & & & & & & & & \\ & & & & & & & & 4 & & & & & & & & & \\ & & & & & & & & & 7 & 0 & 0 & 1 & 0 & & & & \\ & & & & & & & & & & 7 & 0 & 0 & 1 & & & & \\ & & & & & & & & & & & 7 & 0 & 0 & & & & \\ & & & & & & & & & & & & 7 & 0 & 1 & 0 & & \\ & & & & & & & & & & & & & 7 & 0 & 1 & & \\ & & & & & & & & & & & & & & 7 & 0 & & \\ & & & & & & & & & & & & & & & 7 & & \end{array} \right] \quad \square$$

Remark 2.1.7

Our standing assumption throughout this chapter is that the field F is algebraically closed. However, as with the Jordan form, individual results concerning the Weyr form of a matrix $A \in M_n(F)$ still make sense over a general field F provided the characteristic polynomial of A splits into linear factors (i.e., it has a complete set of roots in F). In particular, this applies to nilpotent matrices. \square

Hopefully, by now the reader understands what a Weyr matrix is and is looking forward to finding out, in due course, its properties and applications. We finish off this section by posing a test question, designed to check one's understanding of the Weyr form definition. Please re-read the relevant parts of the section should you fail the test (most unlikely).

Test Question 1. Which of the following matrices are in Weyr form? (The answer is given at the end of the chapter.)

$$\left[\begin{array}{cc|cc} 2 & 0 & 1 & 0 \\ & 2 & 0 & 1 \\ \hline & & 2 & 0 \\ & & & 2 \\ & & & & 2 & 0 & 1 & 0 \\ & & & & & 2 & 0 & 1 \\ & & & & & & 2 & 0 \\ & & & & & & & 2 \end{array} \right] \quad (1)$$

$$\left[\begin{array}{cc|cc} 2 & 0 & 1 & 0 \\ & 2 & 0 & 1 \\ \hline & & 2 & 0 \\ & & & 2 \\ & & & & 3 & 0 & 1 & 0 \\ & & & & & 3 & 0 & 1 \\ & & & & & & 3 & 0 \\ & & & & & & & 3 \end{array} \right] \quad (2)$$

$$\left[\begin{array}{ccc|ccc} 5 & 0 & 0 & 0 & 0 & \\ & 5 & 0 & 1 & 0 & \\ & & 5 & 0 & 1 & \\ \hline & & & 5 & 0 & 1 & 0 \\ & & & & 5 & 0 & 1 \\ & & & & & 5 & 0 \\ & & & & & & 5 \end{array} \right] \quad (3)$$

2.2 EVERY SQUARE MATRIX IS SIMILAR TO A UNIQUE WEYR MATRIX

To qualify for canonical form status, the Weyr form must meet the claim in this section's title. Our goal now is to establish that claim. In the final section of this chapter, we show how the calculations can be done in practice, modulo the old problem of finding the eigenvalues of a matrix. The good news is that the calculations are quite a bit simpler than those for the Jordan form.

In all, we will eventually provide three independent proofs for the existence of the Weyr form: (1) a simple "row operations" proof, (2) a derivation from the Jordan form, and (3) a module-theoretic proof. The first two of these are covered within the present chapter, and the last is developed in Chapter 4.

We begin by recording the following simple observations concerning conjugations by elementary matrices, which will be useful when putting a nilpotent matrix in Weyr form.

Lemma 2.2.1

For $i \neq j$, let $E = E_{ij}(c)$ be the elementary matrix $I + ce_{ij}$, where c is a constant and e_{ij} is the matrix with 1 in the (i, j) position and 0's elsewhere. Then:

- (1) Conjugating a matrix A by E (forming $E^{-1}AE$) has the effect of adding c times the i th column of A to its j th column, and then subtracting c times the j th row of the resulting matrix from its i th row.
- (2) If the i th column of A is zero, then the conjugation in (1) has the same effect on the i th row of A as the corresponding elementary row operation (of subtracting c times the j th row).
- (3) If the first d columns of A are zero, then any elementary row operation (including row swaps) on the first d rows of A can be realized as a conjugation by the corresponding elementary matrix.

Proof

(1) is easily checked, (2) then follows from (1), while in turn (3) follows from (2). \square

Theorem 2.2.2

Every square matrix A over an algebraically closed field F is similar to a matrix in Weyr form.⁵

Proof

By the generalized eigenspace decomposition 1.5.2 and its Corollary 1.5.4, if $\lambda_1, \lambda_2, \dots, \lambda_k$ are the distinct eigenvalues of A , then we know A is similar to $\text{diag}(A_1, A_2, \dots, A_k)$, where each $A_i = \lambda_i I + N_i$ for some nilpotent matrix N_i . Since it is enough to put each N_i in Weyr form, we can suppose our given matrix A is a nilpotent $n \times n$ matrix.

Let d be the nullity of A . Let $V = F^n$ and let \mathcal{B} be the standard basis for V . View A as the matrix relative to \mathcal{B} of the linear transformation of V given by left multiplication by A . Choose a basis for the null space of A and extend this to a basis \mathcal{B}' for V . Under the change of basis, A is transformed to the similar matrix

$$P^{-1}AP = \left[\begin{array}{c|c} 0 & B_2 \\ \hline 0 & A_2 \end{array} \right]$$

5. The second author suggests that this be called “the Weyrisimilitude theorem.”

where $P = [B', B]$ is the change of basis matrix and where the blocked matrix has $d \times d$ and $(n - d) \times (n - d)$ diagonal blocks. (Our labeling of the blocks is to conform to our later algorithm in Section 2.5, where A is denoted by A_1 .) Let $m = n - d$. Observe from Remark 1.2.2 (1) that A_2 is a nilpotent $m \times m$ matrix with $m < n$. Therefore, by induction on n , we can assume A_2 can be put in Weyr form. Let $Q \in GL_m(F)$ be such that $Q^{-1}A_2Q = W$, a nilpotent Weyr matrix. (Here $GL_m(F)$ is the *general linear group*, consisting of all invertible $m \times m$ matrices over F .) Conjugating $P^{-1}AP$ by $\text{diag}(I_d, Q)$ now yields a blocked matrix

$$X = \left[\begin{array}{c|c} 0 & Y \\ \hline 0 & W \end{array} \right]$$

where Y is $d \times m$ and W is an $m \times m$ nilpotent Weyr matrix.

Let $n_1 = d$ and let (n_2, n_3, \dots, n_r) be the Weyr structure of our W . Observe that n_2 is the nullity of W . Since $\text{rank } X \leq \text{rank } W + \text{rank } Y$, we must have $n - n_1 \leq (n - n_1 - n_2) + n_1$ and so $n_1 \geq n_2$. We aim to establish that A can be put in Weyr form with Weyr structure (n_1, n_2, \dots, n_r) . The strategy is simple enough—use the special form of W and elementary row operations on X , in the form of conjugations using Lemma 2.2.1, to transform the first n_1 rows of the unblocked X so as to put X in Weyr form with the proposed Weyr structure. (The rest of X is fine, providing we don't mess it up.) Henceforth, we block our $n \times n$ matrices using the partition $n = n_1 + n_2 + \dots + n_r$, that is, we work with $r \times r$ blocked matrices where the (i, j) block is an ordinary $n_i \times n_j$ matrix. Notice that X , as such a blocked matrix, has the form

$$X = \left[\begin{array}{cccccc} 0 & X_{12} & X_{13} & X_{14} & \cdots & X_{1r} \\ & 0 & I_3 & 0 & \cdots & \vdots \\ & & 0 & I_4 & \cdots & \vdots \\ & & & & \ddots & \vdots \\ & & & & & 0 & I_r \\ & & & & & & 0 \end{array} \right]$$

where, through a slight abuse of notation, I_j denotes the $n_{j-1} \times n_j$ matrix having the $n_j \times n_j$ identity matrix as its upper part followed by $n_{j-1} - n_j$ zero rows, and where $Y = [X_{12}, \dots, X_{1r}]$. By repeated applications of Lemma 2.2.1 (3), utilizing the fact that the first n_1 columns of X are zero, we can use the I_j to successively clear $X_{13}, X_{14}, \dots, X_{1r}$ to zero by conjugating X with suitable elementary matrices that leave the block W unchanged. We are almost there, except we still need to transform X_{12} to I_2 .

Clearly,

$$\begin{aligned}
 n - n_1 &= \text{rank } A \\
 &= \text{rank } X \\
 &= \text{rank } X_{12} + \text{rank } I_3 + \text{rank } I_4 + \cdots + \text{rank } I_r \\
 &= \text{rank } X_{12} + n_3 + n_4 + \cdots + n_r \\
 &= \text{rank } X_{12} + n - n_1 - n_2,
 \end{aligned}$$

which implies $\text{rank } X_{12} = n_2$. That is, X_{12} is an $n_1 \times n_2$ matrix of full column-rank. Therefore, as an $n_1 \times n_2$ matrix, X_{12} can be row-reduced to I_2 . Again by Lemma 2.2.1 (3), since the first n_1 columns of (the transformed) X are zero, we can further conjugate X by suitable elementary $n \times n$ matrices to make the X_{12} block the matrix I_2 , but without affecting the other features of X that we have established. Now we have transformed X (and hence our original matrix A) to a matrix in Weyr form with Weyr structure (n_1, n_2, \dots, n_r) . Our proof is complete. \square

The proof we have given of Theorem 2.2.2 is actually quite constructive, involving two phases: (1) finding bases for the null spaces of decreasingly smaller matrices and then extending them to bases for the underlying spaces, and (2) using elementary row operations to either put a submatrix in reduced row-echelon form or to clear out a submatrix using an identity submatrix from lower down. We will set all this out as an algorithm in the last section.

We next establish the uniqueness of the Weyr form, to the same degree of uniqueness that one has for the Jordan form—unique to within permutation of basic blocks. Fix an $n \times n$ matrix A . Suppose we have two Weyr matrices

$$W = \begin{bmatrix} W_1 & & & \\ & W_2 & & \\ & & \ddots & \\ & & & W_k \end{bmatrix}, \quad W' = \begin{bmatrix} W'_1 & & & \\ & W'_2 & & \\ & & \ddots & \\ & & & W'_k \end{bmatrix}$$

and each is similar to A . Here W_1, W_2, \dots, W_k are basic Weyr matrices corresponding to the different eigenvalues of A . The same applies to W' but the order of the eigenvalues may not match. By permuting the basic blocks of W' , we can suppose the eigenvalue corresponding to the i th basic block in each of W and W' is λ_i where $\lambda_1, \lambda_2, \dots, \lambda_k$ are the distinct eigenvalues of A in some order. Now we want to show $W = W'$. By Proposition 1.6.2, we know W_i is similar to W'_i for $i = 1, 2, \dots, k$. Therefore, to establish uniqueness of the Weyr form, it suffices to show that two similar *basic* Weyr matrices with eigenvalue λ must have the same Weyr structure (for then the two matrices are equal). This we do in the following proposition. There we show that the

Weyr structure of a basic Weyr matrix W with eigenvalue λ is completely determined by the nullities of the powers of $W - \lambda I$. This is analogous to the situation for determining the Jordan structure of a matrix having λ as its single eigenvalue (see Corollary 2.4.6). However, those familiar with a direct proof of this connection with the Jordan structure will notice that the nullity connections for basic Weyr matrices are much cleaner and easier to establish.⁶

Proposition 2.2.3

If W is a basic Weyr matrix with eigenvalue λ and Weyr structure (n_1, n_2, \dots, n_r) , then

$$\begin{aligned} r &= \text{nilpotent index of } W - \lambda I, \\ n_1 &= \text{nullity}(W - \lambda I), \\ n_i &= \text{nullity}(W - \lambda I)^i - \text{nullity}(W - \lambda I)^{i-1} \text{ for } i = 2, \dots, r. \end{aligned}$$

Consequently, two similar basic Weyr matrices are equal.

Proof

Let $N = W - \lambda I$ and view N and its powers as $r \times r$ blocked matrices relative to the partition $n = n_1 + n_2 + \dots + n_r$. Let I_j denote an appropriately sized matrix with n_j columns and having the $n_j \times n_j$ identity matrix as its upper part, followed by zero rows. (Our abuse of notation here has progressed one more stage—the number of rows of I_j depends on its block location within the block matrix! That shouldn't cause confusion. Just remember as we move up the j th column of blocks, the number of columns in the blocks remains steady at n_j , but the number of rows in the blocks increases.) We have

$$N = \begin{bmatrix} 0 & I_2 & 0 & 0 & \cdots & 0 & 0 \\ & 0 & I_3 & 0 & \cdots & 0 & 0 \\ & & 0 & & & & \vdots \\ & & & \ddots & & & \\ & & & & I_{r-1} & 0 \\ & & & & 0 & I_r \\ & & & & & 0 \end{bmatrix}$$

6. Many find the proof of the uniqueness of the Jordan form quite a challenge—even very good students.

and

$$N^i = \begin{bmatrix} 0 & \cdots & 0 & I_{i+1} & 0 & \cdots & 0 \\ & & & 0 & I_{i+2} & 0 & \vdots \\ & & & & \ddots & \ddots & \\ & & & & & 0 & I_r \\ & & & & & & 0 \\ & & & & & & \vdots \\ & & & & & & 0 \end{bmatrix}$$

for $i = 1, \dots, r-1$. Clearly N is nilpotent of index r . Now for $i = 1, \dots, r-1$ we have $\text{rank } N^i = n_{i+1} + n_{i+2} + \cdots + n_r$, giving $n_i = \text{rank } N^{i-1} - \text{rank } N^i = \text{nullity } N^i - \text{nullity } N^{i-1}$. The last expression equals n_r also for $i = r$. The final statement of the proposition follows from the simple fact that similar matrices have the same nullity, together with the observation that if X and Y are similar, then so are $(X - \lambda I)^i$ and $(Y - \lambda I)^i$ (under the same similarity transformation—think of images under an automorphism). \square

Combining Theorem 2.2.2 and Proposition 2.2.3 gives us our principal result for this section.

Theorem 2.2.4

*To within permutation of basic Weyr blocks, each square matrix A over an algebraically closed field is similar to a unique Weyr matrix W . The matrix W is called the **Weyr (canonical) form** of A .*

In view of Theorem 2.2.4, it makes sense to define the **Weyr structure of a matrix A associated with an eigenvalue λ** to be the Weyr structure of the basic Weyr block W with eigenvalue λ that occurs in the unique Weyr form of A . If we let

$$\omega_i = \text{nullity}(A - \lambda I)^i - \text{nullity}(A - \lambda I)^{i-1} \quad \text{for } i = 1, 2, 3, \dots$$

then the sequence $\omega_1, \omega_2, \omega_3, \dots$ has been historically referred to as the **Weyr characteristic of A associated with the eigenvalue λ** . In our terminology,⁷ the initial finite sequence $(\omega_1, \omega_2, \dots, \omega_r)$ of nonzero terms of the Weyr characteristic agrees with the Weyr structure of A associated with λ because of Proposition 2.2.3. Here r can be determined as the least positive integer for

7. The corresponding Jordan form term is historically referred to as the *Segre characteristic*, with no mention of Jordan. For the sake of consistency, we choose “Jordan structure” over “Segre characteristic,” and “Weyr structure” over “Weyr characteristic.”

which $\text{nullity}(A - \lambda I)^r = \text{nullity}(A - \lambda I)^{r+1}$. It is important to bear in mind, however, a conceptual distinction between “Weyr characteristic” and “Weyr structure”: the Weyr characteristic of a matrix A is defined independently of the Weyr form. On the other hand, the Weyr structure of A describes the shape of the Weyr form of A .

The following observation, immediate from Proposition 2.2.3, is used implicitly many times throughout our book.

Proposition 2.2.5

Suppose F is an algebraically closed field. Two matrices $A, B \in M_n(F)$ are similar if and only if they have the same eigenvalues and the same Weyr characteristics (or Weyr structures) associated with these eigenvalues.

Remarks 2.2.6

- (1) The Weyr characteristic of a nilpotent matrix A can be calculated in an efficient way using our next corollary, and this doesn’t require completely putting A in Weyr form (see also our later Remark 2.5.1). For complex matrices, the Weyr characteristic can be calculated using only unitary similarities that put the matrix in a certain block upper triangular form (see later Remark 2.5.2). Historically, these developments have been something of a mixed blessing, because some have mistakenly concluded that therefore the Weyr form itself is just an optional extra in linear algebra. It may also help explain why the Weyr characteristic is a better-known concept than the Weyr form.
- (2) There are many interesting applications of the Weyr characteristic, and our downplaying of the concept to a few remarks does not do it justice. However, our book studies in depth the Weyr form, and space limitations have prevented us from covering the Weyr characteristic in depth (outside, of course, how it relates to the Weyr structure of a matrix). To see how the Weyr characteristic relates to the singular graph of an M -matrix, see the papers by Richman (1978–79), Richman and Schneider (1978), and Hershkowitz and Schneider (1989, 1991a and 1991b). See Shapiro’s article for a hint of applying the Weyr characteristic to the important problem of computing the Jordan form of a complex matrix in a stable manner. For a lovely application of the Weyr characteristic to a “pure” linear algebra problem, see the 2009 article “The Jordan Forms of AB and BA ” by Lippert and Strang. Here the authors cleverly utilize the fact that the Weyr characteristic and Jordan structure of a nilpotent matrix are dual partitions. (See our later Theorem 2.4.1.) □

In computing the Weyr structure of a given nilpotent matrix A , one can avoid computing the powers of A and their nullities directly, and then invoking Proposition 2.2.3, by a recursive use of the following corollary to the proof of Theorem 2.2.2:

Corollary 2.2.7

Suppose $A \in M_n(F)$ is nilpotent and has nullity d . Let $m = n - d$ and suppose A is similar to the block matrix

$$\left[\begin{array}{c|c} 0 & B \\ \hline 0 & C \end{array} \right]$$

where the top left block is $d \times d$ and C is $m \times m$. If (n_2, n_3, \dots, n_r) is the Weyr structure of the nilpotent matrix C , then the Weyr structure of A is $(d, n_2, n_3, \dots, n_r)$.

Although of limited practical use for large matrices, our next result uses Proposition 2.2.3 to give a complete test for when two matrices are similar, without having to produce a similarity transformation. Its proof is a good illustration of the power of canonical forms—in fact, it is not clear how one could establish the result without a canonical form. Our weapon of choice is the Weyr form, whose 2.2.3 analogue we give in Corollary 2.4.6.

Proposition 2.2.8

Two $n \times n$ matrices A and B over an algebraically closed field are similar if and only if they have the same distinct eigenvalues and

$$\text{nullity}(A - \lambda I)^j = \text{nullity}(B - \lambda I)^j$$

for each eigenvalue λ and for $j = 1, 2, \dots, n$.

Proof

Clearly, these conditions are necessary for A and B to be similar. Now suppose the conditions hold for A and B . We show A and B are similar by arguing that they have the same Weyr form.

Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be the distinct eigenvalues of A and B . By Corollary 1.5.4 we may assume

$$A = \text{diag}(A_1, A_2, \dots, A_k),$$

$$B = \text{diag}(B_1, B_2, \dots, B_k),$$

where A_i and B_i have λ_i as their only eigenvalue. It is enough to show A_i and B_i have the same Weyr structure. This we establish for $i = 1$; the general case is entirely similar. To ease notation, set $\lambda = \lambda_1$, $X = A_1$, and $Y = B_1$.

Since $(A - \lambda I)^j = \text{diag}((X - \lambda I)^j, *, *, \dots, *)$ where the $*$ blocks are invertible (for $i \geq 2$, the i th block has the nonzero $(\lambda_i - \lambda)^j$ as its single

eigenvalue), we see that $\text{nullity}(A - \lambda I)^j = \text{nullity}(X - \lambda I)^j$ for all j . Similarly, $\text{nullity}(B - \lambda I)^j = \text{nullity}(Y - \lambda I)^j$. Hence, by our hypotheses, we have

$$\text{nullity}(X - \lambda I)^j = \text{nullity}(Y - \lambda I)^j \text{ for } j = 1, 2, \dots, n.$$

In particular, since $X - \lambda I$ and $Y - \lambda I$ are nilpotent matrices, this implies X and Y must be of the same size. (For an $m \times m$ nilpotent matrix N , we know $N^m = 0$ from the Cayley-Hamilton theorem, whence N^m and higher powers all have nullity m .) It now follows from Proposition 2.2.3 that X and Y have the same Weyr structure, as we sought to establish. \square

Notice that the proposition gives a nice “one line” proof of the fact that a matrix $A \in M_n(F)$ is similar to its transpose $B = A^T$.⁸ A square matrix and its transpose have the same eigenvalues (since $\det(xI - A) = \det(xI - A)^T = \det(xI - A^T)$), and the same nullity (because row rank equals column rank). Moreover, $(A^T - \lambda I)^i = ((A - \lambda I)^i)^T$ (whence the powers $(A - \lambda I)^i$ and $(A^T - \lambda I)^i$ have the same nullity).

As a source of amusement, the reader may wish to check the claims concerning similarity of the matrices A, B, C in the following example (it is almost as simple as “ABC”).

Example 2.2.9

The following three 5×5 matrices each have $\lambda = 2$ as their sole eigenvalue:

$$A = \begin{bmatrix} 4 & -6 & 1 & 2 & 0 \\ 1 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 1 \\ 1 & -3 & -1 & 3 & 0 \\ 1 & -3 & 0 & 1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 1 & 0 & 1 & 1 \\ -2 & 0 & -1 & -2 & -2 \\ 1 & 0 & 2 & -1 & 0 \\ 1 & 1 & 0 & 3 & 1 \\ 0 & 0 & 1 & 0 & 2 \end{bmatrix},$$

$$C = \begin{bmatrix} 2 & 0 & 1 & 0 & 0 \\ 2 & 3 & -1 & 0 & -1 \\ 0 & 0 & 2 & 1 & 0 \\ -2 & -1 & 1 & 2 & 1 \\ 2 & 1 & 1 & -1 & 1 \end{bmatrix}.$$

8. Although this result is not unexpected, a direct proof (not involving canonical forms) is surprisingly difficult. A common mistake is to look for a fixed invertible $n \times n$ matrix C , for example a permutation matrix, such that $C^{-1}AC = A^T$ for all $A \in M_n(F)$. When $n > 1$, such a matrix can't exist! Otherwise, transposing matrices would be an algebra automorphism. But in general $(XY)^T = Y^T X^T \neq X^T Y^T$. Thus, the similarity transformation must be tailored to the individual matrix A .

Subject to the accuracy of the information contained in lines 2–4 of each of the following tables, we can deduce from Proposition 2.2.8 that A is similar to B , but not to C . The last line, giving the Weyr structure, follows from Proposition 2.2.3.

index of $A - 2I$	3		index of $B - 2I$	3	
nullity($A - 2I$)	2		nullity($B - 2I$)	2	
nullity($A - 2I$) ²	4		nullity($B - 2I$) ²	4	
nullity($A - 2I$) ³	5	,	nullity($B - 2I$) ³	5	,
nullity($A - 2I$) ⁴	5		nullity($B - 2I$) ⁴	5	
Weyr structure of A	(2, 2, 1)		Weyr structure of B	(2, 2, 1)	

index of $C - 2I$	4
nullity($C - 2I$)	2
nullity($C - 2I$) ²	3
nullity($C - 2I$) ³	4
nullity($C - 2I$) ⁴	5
Weyr structure of C	(2, 1, 1, 1)

□

2.3 SIMULTANEOUS TRIANGULARIZATION

Upper triangular matrices are simpler to work with in many situations, in particular for deciding commuting relations. Moreover, it is well known that any set of commuting $n \times n$ matrices over an algebraically closed field F can be simultaneously triangularized. We will give the short proof of this and then extend the result by showing that, in addition, we can require that the first matrix actually finishes up in Weyr form. *This is not always possible with the Jordan form.* See Example 2.3.6.

If we have fixed upon a particular matrix $A \in M_n(F)$, then any other matrix $B \in M_n(F)$ that commutes with A is said to *centralize* A . The set of all matrices B that centralize A is called the **centralizer** of A , and we denote it by $\mathcal{C}(A)$. This is fairly standard terminology,⁹ and is consistent with the use of the word “centralize” in, say, group theory. Chapter 3 is devoted entirely to centralizers. However, in this section we have an immediate need for the description of

9. A few authors avoid the term “centralize” altogether. But then they are faced with using the outdated term “commutant” for the centralizer set $\mathcal{C}(A)$.

the centralizer of a nilpotent Weyr matrix, in order to establish the extended triangularization property. So we will give that description next and then return to the full study in Chapter 3. If the authors had to pick out just one feature of the Weyr form that makes it so useful for our later applications, better than the Jordan form, they would go for the description of the centralizer of a nilpotent Weyr matrix.

In the calculations that follow (not just in this chapter), the reader can compute a particular product involving a nilpotent Weyr matrix W by simply multiplying out the matrices. However, there is often a much better and quicker “visual” way. This involves watching how W shifts columns under right multiplication, and shifts rows under left multiplication.¹⁰ The shifting is a blocked matrix version of the behavior that is exemplified in the following two products involving the 5×5 basic Jordan nilpotent matrix J and the “alphabet” matrix A (watch what is happening to rows and columns):

$$AJ = \begin{bmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & p \\ q & r & s & t & u \\ v & w & x & y & z \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & a & b & c & d \\ 0 & f & g & h & i \\ 0 & k & l & m & n \\ 0 & q & r & s & t \\ 0 & v & w & x & y \end{bmatrix},$$

$$JA = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & p \\ q & r & s & t & u \\ v & w & x & y & z \end{bmatrix} = \begin{bmatrix} f & g & h & i & j \\ k & l & m & n & p \\ q & r & s & t & u \\ v & w & x & y & z \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The “shifting” under multiplication by a nilpotent Weyr matrix is one of the most suggestive features of the Weyr form. And at the risk of appearing as cheeky as a kea,¹¹ we claim the Jordan form shifting, both in its suggestiveness and utility, does not hold a candle to the Weyr shifting. For instance, suppose J and W are nilpotent Jordan and Weyr matrices, respectively, and to make for

10. In a product AB of square matrices, one has a choice of two views of the resulting matrix: (i) how A has affected the *row* pattern of B or (ii) how B has affected the *column* pattern of A . With the former, we can pick out the pattern by observing how the rows of A are related to those of the identity matrix I (because $A = AI$). With the latter, we see how the columns of B can be formed from those of I (since $B = IB$).

11. The kea is a large, New Zealand alpine parrot, known for its intelligence and bold opportunism. A kea once made off with a tourist’s passport, the passport never to be seen again. (There have been no reports of the kea presenting himself at an overseas port.)

a level playing field, assume they have the same structure, say homogeneous $(3, 3, 3)$. In a product AJ or AW , it is natural to block all matrices according to diagonal block sizes $3, 3, 3$. For a general A this means writing

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix},$$

where the A_{ij} are 3×3 matrices. Now

$$AW = \begin{bmatrix} 0 & A_{11} & A_{12} \\ 0 & A_{21} & A_{22} \\ 0 & A_{31} & A_{32} \end{bmatrix}.$$

But it is impossible to express AJ as a blocked matrix whose block entries are in terms of only the A_{ij} . This is because J shifts locally, not globally. For instance, AJ has for its $(1, 1)$ block the 3×3 matrix obtained by shifting the columns of A_{11} .

Since we will have many occasions to refer back to the shifting, we record it in the following remark.

Remark 2.3.1

An $n \times n$ basic Jordan nilpotent matrix acts under **right** multiplication on a general matrix X by shifting the columns of X one step to the right, introducing a zero first column and killing the last column. A general nilpotent Jordan matrix does this shifting only “locally,” acting on individual blocks. However, a general nilpotent Weyr matrix W does this shifting (under right multiplication) “globally” on blocked matrices, shifting blocks to the right but with one proviso: Suppose the Weyr structure of W is (n_1, n_2, \dots, n_r) , and we view X as an $r \times r$ blocked matrix (X_{ij}) relative to the partition $n = n_1 + n_2 + \dots + n_r$ (so the block X_{ij} is $n_i \times n_j$). Then under right multiplication, W can’t faithfully shift the j th column of blocks of X to the $(j + 1)$ th column if $n_j > n_{j+1}$. (Remember $n_j \geq n_{j+1}$ always.) In this case only the first n_{j+1} columns of X_{ij} are shifted, and the remaining $n_j - n_{j+1}$ are deleted. **Left** multiplication by W has a similar shift effect on the rows of blocks of X , shifting from the bottom upwards, and appending $n_i - n_{i+1}$ zero rows to $X_{(i+1)j}$ whenever $n_i > n_{i+1}$. \square

We now illustrate this shifting and appending action with an example.

Example 2.3.2

In the multiplication below, the nilpotent Weyr matrix W has the Weyr structure $(3, 2, 2)$, and W centralizes the leftmost matrix X (the case of principal interest

later). We have for the product XW ,

$$\left[\begin{array}{ccc|ccc} a & b & e & h & i & l & m \\ c & d & f & j & k & n & p \\ 0 & 0 & g & 0 & 0 & q & r \\ \hline & & & a & b & h & i \\ & & & c & d & j & k \\ & & & & & a & b \\ & & & & & c & d \end{array} \right] \left[\begin{array}{ccc|cc} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline & & & 0 & 0 & 1 & 0 \\ & & & 0 & 0 & 0 & 1 \\ & & & & & 0 & 0 \\ & & & & & 0 & 0 \end{array} \right] = \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & a & b & h & i \\ 0 & 0 & 0 & c & d & j & k \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline & & & 0 & 0 & a & b \\ & & & 0 & 0 & c & d \\ & & & & & 0 & 0 \\ & & & & & 0 & 0 \end{array} \right].$$

We observe that the right-hand matrix can be obtained by both the column shifting and row shifting on X described in the remark (because $XW = WX$). And isn't it so easy to do! \square

The centralizer of an $n \times n$ basic Jordan matrix J is easily calculated (as in Chapter 3) to be the set of upper triangular matrices $K = (k_{ij})$ ¹² for which entries in the same superdiagonal (including the diagonal) are equal:¹³

$$K = \begin{bmatrix} a & b & c & \dots & & \\ & a & b & c & \dots & \\ & & a & b & c & \dots \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & \ddots \end{bmatrix}$$

That is, $k_{ij} = 0$ for $i > j$ and $k_{ij} = k_{i+1,j+1}$ for $1 \leq i \leq j \leq n-1$. (So the centralizer is the subalgebra generated by J , which is to be expected by Proposition 1.1.2.) For an $n \times n$ basic Weyr matrix W , the centralizer is a little more complicated but nevertheless has a similar description in terms of block upper triangular matrices, if we weaken the requirement that the (i, j) and $(i+1, j+1)$ blocks be “equal” when the second block is strictly smaller. Observe the pattern in the matrix $K = X$ of Example 2.3.2 above. That is typical.

Proposition 2.3.3

Let W be an $n \times n$ basic Weyr matrix with the Weyr structure (n_1, \dots, n_r) , $r \geq 2$. Let K be an $n \times n$ matrix, blocked according to the partition $n = n_1 + n_2 + \dots + n_r$,

12. As a general rule, we use the letter K to denote a centralizing matrix.

13. In other parlance, K is an upper triangular “Toeplitz” matrix.

and let K_{ij} denote its (i, j) block (an $n_i \times n_j$ matrix) for $i, j = 1, \dots, r$. Then W and K commute if and only if K is a block upper triangular matrix for which

$$K_{ij} = \begin{bmatrix} K_{i+1,j+1} & * \\ 0 & * \end{bmatrix} \text{ for } 1 \leq i \leq j \leq r-1.$$

Here we have written K_{ij} as a blocked matrix where the zero block is $(n_i - n_{i+1}) \times n_{j+1}$. The asterisk entries (*) indicate that there are no restrictions on the entries in that part of the matrix. (The column of asterisks disappears if $n_j = n_{j+1}$, and the $[0 \ *]$ row disappears if $n_i = n_{i+1}$.)

Proof

By subtracting the diagonal of W , we can assume W is nilpotent without changing its centralizer. For $j = 2, \dots, r$, let I_j denote the $n_{j-1} \times n_j$ matrix having the $n_j \times n_j$ identity matrix as its upper part followed by $n_{j-1} - n_j$ zero rows. Then as a blocked matrix

$$W = \begin{bmatrix} 0 & I_2 & & & \\ & 0 & I_3 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 0 & I_r \\ & & & & 0 & 0 \end{bmatrix}.$$

Suppose K commutes with W . By examining the first column of blocks in $KW = WK$, we obtain

$$K_{21} = K_{31} = \dots = K_{r1} = 0. \quad (4)$$

(For instance, one can use the shifting effects described in Remark 2.3.1.) It also follows easily from the equation $KW = WK$ that

$$K_{ij}I_{j+1} = I_{i+1}K_{i+1,j+1} \text{ for } 1 \leq i, j \leq r-1. \quad (5)$$

Next observe that (5) implies the form connecting K_{ij} and $K_{i+1,j+1}$, namely

$$K_{ij} = \begin{bmatrix} K_{i+1,j+1} & * \\ 0 & * \end{bmatrix},$$

but for all i, j with $1 \leq i, j \leq r-1$. However, this latter property, in tandem with (4), forces $K_{ij} = 0$ for all $i > j$, thus completing the “only if” part of the proof.

For the converse, assume K satisfies the conditions of the proposition. A simple calculation then shows

$$K_{ij}I_{j+1} = I_{i+1}K_{i+1,j+1} \text{ for } 1 \leq i \leq j \leq r-1.$$

Therefore the $(i, j+1)$ block entries of KW and WK agree within the stated range of i and j , that is, on all blocks above the diagonal. Inasmuch as K is block upper triangular, Remark 2.3.1 implies that all the diagonal block entries of KW and WK must be zero. Because both products are block upper triangular, all block entries of KW and WK must agree. \square

Proposition 2.3.4

Let A_1, A_2, \dots, A_k be commuting $n \times n$ matrices over an algebraically closed field F . Then A_1, A_2, \dots, A_k can be simultaneously triangularized. That is, for some invertible matrix $C \in M_n(F)$, each $C^{-1}A_iC$ is an upper triangular matrix for $i = 1, 2, \dots, k$.

Proof

Let $V = F^n$, let \mathcal{B} be the standard basis, and think of each matrix A_j as the matrix relative to \mathcal{B} of the left multiplication map of V by A_j . Clearly we can suppose that not all the matrices in our list are scalar matrices, and that, after reordering, A_1 is nonscalar. Since F is algebraically closed, we can choose an eigenvalue λ of A_1 . Let $U = \ker(\lambda I - A_1)$ be the corresponding eigenspace. By Proposition 1.3.1, U is invariant under each A_j because A_j commutes with $\lambda I - A_1$. Let $m = \dim U$. Since A_1 is not a scalar matrix, we have $1 \leq m < n$. Choose a basis \mathcal{B}' for V that extends some basis for U . The matrix of the multiplication map by A_j relative to \mathcal{B}' takes the form

$$P^{-1}A_jP = \begin{bmatrix} B_j & C_j \\ 0 & D_j \end{bmatrix}$$

where $P = [\mathcal{B}', \mathcal{B}]$ is the change of basis matrix, B_j is $m \times m$, and D_j is $(n-m) \times (n-m)$. Since the $P^{-1}A_jP$ commute, so do the B_j , and the D_j for $j = 1, 2, \dots, k$. By induction on n , we can assume that the B_j can be simultaneously triangularized, using, say, conjugation by $R \in GL_m(F)$, and that the D_j can be simultaneously triangularized, say, using $S \in GL_{n-m}(F)$. Then conjugation by

$$Q = \begin{bmatrix} R & 0 \\ 0 & S \end{bmatrix}$$

makes all the $P^{-1}A_jP$ upper triangular. Thus, conjugating A_1, A_2, \dots, A_k by $C = PQ$ yields our desired result.¹⁴ \square

Theorem 2.3.5

Let A_1, A_2, \dots, A_k be commuting $n \times n$ matrices over an algebraically closed field F . Then there is a similarity transformation that puts A_1 in Weyr form and simultaneously puts A_2, \dots, A_k in upper triangular form.

Proof

Using the Corollary 1.5.4 to the generalized eigenspace decomposition 1.5.2, we can simultaneously conjugate A_1, A_2, \dots, A_k to put A_1 in block diagonal form with each block having a single eigenvalue and different blocks having different eigenvalues. Because of commutativity and the distinctness of the eigenvalues in different blocks, there is a matching block diagonal splitting of the other A_i but without eigenvalue restrictions. (This argument is spelled out more fully in Proposition 3.1.1 for those readers not already familiar with it.) Hence, we can reduce to the case where A_1 has only a single eigenvalue.

By Theorem 2.2.2 (and another simultaneous conjugation), we can assume the matrices A_1, A_2, \dots, A_k are commuting, with A_1 a basic Weyr matrix, say, of Weyr structure (n_1, n_2, \dots, n_r) . Applying Proposition 2.3.3 to $W = A_1$, we see that each A_i is a block upper triangular matrix with respect to the A_1 block structure. We complete the proof by inductively constructing an invertible block diagonal matrix $C = \text{diag}(C_1, \dots, C_r)$ of block structure (n_1, n_2, \dots, n_r) such that

- (i) C centralizes A_1 , and
- (ii) C conjugates A_2, \dots, A_k simultaneously to (properly, not just block) upper triangular matrices.

Then conjugating A_1, A_2, \dots, A_k by C leaves A_1 in Weyr form and simultaneously makes A_2, \dots, A_k upper triangular. We construct the C_i recursively in the order C_r, C_{r-1}, \dots, C_1 .

The (r, r) diagonal blocks of A_2, \dots, A_k commute, so by Proposition 2.3.4 there is an invertible $n_r \times n_r$ matrix C_r that simultaneously conjugates these blocks to upper triangular matrices. Suppose now we have constructed C_i for some $i > 1$. Here is how we construct C_{i-1} . If $n_{i-1} = n_i$, we set $C_{i-1} = C_i$. Suppose $n_{i-1} > n_i$. Since A_2, \dots, A_k centralize A_1 , by Proposition 2.3.3 the $(i-1, i-1)$ block of A_j has the form

$$\begin{bmatrix} Y_j & * \\ 0 & Z_j \end{bmatrix} \quad \text{for } j = 2, \dots, k,$$

14. Exactly the same proof works for an infinite set of commuting matrices. Alternatively, the subalgebra generated by such matrices, being finite-dimensional, must be finitely generated, and once the commuting generators are triangularized, so are all matrices in the subalgebra.

where Y_j is the (i, i) block of A_j and Z_j is an $(n_{i-1} - n_i) \times (n_{i-1} - n_i)$ matrix. (The Y 's and Z 's also depend on i , but view i as fixed in this discussion, to avoid a double indexing.) The Z_j commute because A_2, \dots, A_k do. Choose an invertible $(n_{i-1} - n_i) \times (n_{i-1} - n_i)$ matrix D_{i-1} that simultaneously conjugates Z_2, \dots, Z_k to upper triangular matrices. Now set

$$C_{i-1} = \begin{bmatrix} C_i & 0 \\ 0 & D_{i-1} \end{bmatrix}.$$

This completes the construction of $C = \text{diag}(C_1, \dots, C_r)$. But why do (i) and (ii) hold? For (i), our construction guarantees that C centralizes A_1 by Proposition 2.3.3. For (ii), first check inductively that for each $i = r, r-1, \dots, 1$, C_i conjugates the (i, i) blocks of A_2, A_3, \dots, A_k to upper triangular $n_i \times n_i$ matrices. Next, observe that for a general $r \times r$ block upper triangular matrix $X = (X_{ij})$ (of the same block structure as C), we have

$$C^{-1}XC = \begin{bmatrix} C_1^{-1}X_{11}C_1 & * & \cdots & * \\ 0 & C_2^{-1}X_{22}C_2 & * & * \\ 0 & 0 & \ddots & * \\ \vdots & & & \\ 0 & 0 & \cdots & C_r^{-1}X_{rr}C_r \end{bmatrix}.$$

In particular, letting X range over A_2, A_3, \dots, A_k , we see that property (ii) follows. \square

Example 2.3.6

Let

$$A_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then A_1 and A_2 commute, but it is not possible under a similarity transformation to put A_1 in Jordan form and simultaneously put A_2 in upper triangular form. For suppose $C \in M_4(F)$ is an invertible matrix such that $C^{-1}A_1C$ is in Jordan form

and $C^{-1}A_2C = T$ is upper triangular. Then, since A_1 is already in Jordan form, $C^{-1}A_1C = A_1$ and so C centralizes A_1 . Hence, C has the form

$$C = \begin{bmatrix} a & b & c & d \\ 0 & a & 0 & c \\ e & f & g & h \\ 0 & e & 0 & g \end{bmatrix}.$$

From $A_2C = CT$, we obtain

$$\begin{bmatrix} e & f & g & h \\ 0 & e & 0 & g \\ 0 & a & 0 & c \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ 0 & a & 0 & c \\ e & f & g & h \\ 0 & e & 0 & g \end{bmatrix} \begin{bmatrix} x & y & * & * \\ 0 & z & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{bmatrix}$$

for some x, y, z . Equating entries for the first column of each side gives $e = 0$, then equating entries for the second column gives $a = 0$. But now C has zero first column, contradicting the invertibility of C . \square

Remark 2.3.7

The alert reader may have noticed that in the previous example, since A_1 is a polynomial in A_2 ($A_1 = A_2^2$), it is possible to simultaneously put A_2 in Jordan form and have A_1 be upper triangular. (In fact, here one only need conjugate A_1 and A_2 by the permutation matrix obtained by swapping columns 2 and 3 of I .) So that raises the question of whether, given a finite collection of commuting matrices, it is possible to put at least one of them in Jordan form and have the rest upper triangular. The answer is still “no,” but the counterexamples are harder to verify. The interested reader can check that the following two commuting 8×8 matrices give a counterexample. We present the matrices as blocked matrices with 4×4 blocks, expressed in terms of the two matrices A_1 and A_2 in Example 2.3.6:

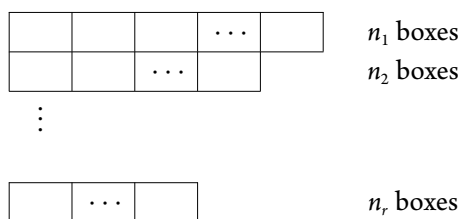
$$\left[\begin{array}{c|c} A_1 & 0 \\ \hline 0 & I + A_2 \end{array} \right], \quad \left[\begin{array}{c|c} A_2 & 0 \\ \hline 0 & I + A_1 \end{array} \right]$$

The reason for introducing the identity matrix in the second diagonal block of each is so that the first and second diagonal blocks now have different eigenvalues. Thus, a centralizing matrix of either must be block diagonal. \square

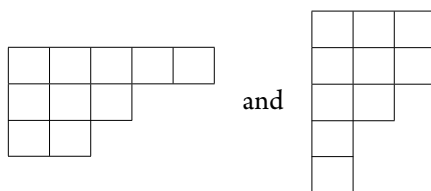
2.4 THE DUALITY BETWEEN THE JORDAN AND WEYR FORMS

There is a lovely duality between the Jordan forms and Weyr forms of $n \times n$ nilpotent matrices, which we will establish in this section. As a by-product, we obtain our second way of establishing the existence and uniqueness of the Weyr form, by appealing to the same aspects of the Jordan form. Equally, this duality combined with Theorem 2.2.4 now provides a relatively short “row operations” proof for the Jordan form. Perhaps the reader has foreseen the duality when we first motivated the Weyr form in Section 2.1, by reordering bases. The duality enables one to mentally flip back and forth between the two forms and decide which form may be the better in a particular circumstance (e.g., notationally or computationally).

The duality involves “dual” (“conjugate” or “transpose”) partitions of n . Recall that each partition (n_1, n_2, \dots, n_r) of n determines a *Young tableau* (or *Ferrer’s diagram*):



“Transposing” the tableau (writing its columns as rows) gives a Young tableau that corresponds to the **dual** partition (m_1, m_2, \dots, m_s) of (n_1, n_2, \dots, n_r) . Thus, in terms of the original tableau, m_1 is the number of first position boxes ($= r$), m_2 is the number of second position boxes, and so on down to m_s being the number of n_1^{th} position boxes. (Thus, $s = n_1$.) For instance,



are the tableaux corresponding to the dual partitions $(5, 3, 2)$ and $(3, 3, 2, 1, 1)$ of 10.

Theorem 2.4.1

The Weyr and Jordan structures of a nilpotent $n \times n$ matrix A (more generally, a matrix with a single eigenvalue) are dual partitions of n . Moreover, the Weyr

form and Jordan form of a square matrix are conjugate under a permutation transformation.

Proof

The argument amounts to just formalizing our initial discussion in Section 2.1 involving reordering bases, by going from a row order to a column order. There we started with the Jordan form and derived the Weyr form. This time, we turn things around to emphasize the duality in the other direction. Thus, we assume that A is a nilpotent Weyr matrix, say with Weyr structure (n_1, n_2, \dots, n_r) . View A as the matrix of a transformation $T: F^n \rightarrow F^n$ relative to an ordered basis $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$. Write $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_r$ where $\mathcal{B}_1 = \{v_1, \dots, v_{n_1}\}$ consists of the first n_1 basis vectors, \mathcal{B}_2 the next n_2 basis vectors, and so on. From the form of A , the action of T on \mathcal{B} is to annihilate \mathcal{B}_1 and then shift (in order) the n_i vectors in \mathcal{B}_i to the corresponding first n_i vectors in \mathcal{B}_{i-1} for $i = 2, \dots, r$. Now reorder the basis \mathcal{B} as $\mathcal{B}' = \mathcal{B}'_1 \cup \mathcal{B}'_2 \cup \dots \cup \mathcal{B}'_s$ where \mathcal{B}'_1 consists of the first members of $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_r$ (in the ordering of \mathcal{B}), while \mathcal{B}'_2 consists of the second members of $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_r$ (no contribution from those \mathcal{B}_i with $|\mathcal{B}_i| = 1$), and so on down to \mathcal{B}'_s which consists of all the last members (of those \mathcal{B}_i with $|\mathcal{B}_i| = n_1$). We have the following Young diagram¹⁵ in which the boxes contain the basis vectors of \mathcal{B} distributed in its rows and the basis vectors of \mathcal{B}' distributed in its columns.

	\mathcal{B}'_1	\mathcal{B}'_2	\mathcal{B}'_3		\mathcal{B}'_s
\mathcal{B}_1	v_1	v_2	v_3	\dots	v_{n_1}
\mathcal{B}_2			\dots		
\vdots					
\vdots					
\mathcal{B}_r		\dots	v_n		

Using our earlier observation on how T acts on vectors in \mathcal{B} , we see that T acts cyclically on each \mathcal{B}'_i , by shifting each vector to its predecessor and then annihilating the first. Thus, the matrix J of T relative to \mathcal{B}' is the Jordan form of A , whence the Jordan structure of A is (m_1, m_2, \dots, m_s) where $m_i = |\mathcal{B}'_i|$ for $i = 1, 2, \dots, s$. Therefore, from the above diagram, the Weyr and Jordan structures are dual partitions. Moreover, $J = P^{-1}AP$ where $P = [\mathcal{B}', \mathcal{B}]$ is the change of basis matrix, which is just the permutation matrix corresponding to the reordering of the basis vectors. (One permutes the rows of the identity matrix.)

In the general case in which A is a direct sum of basic Weyr matrices corresponding to the distinct eigenvalues of A , one does this permutation

15. Once objects are placed in the boxes of a Young tableau, it becomes a Young diagram.

conjugation on each of the basic Weyr matrices and obtains the Jordan form. \square

Remarks 2.4.2

- (1) As a result of the duality in Theorem 2.4.1, in a theoretical sense, anything that can be done with the Weyr form can be done with the Jordan form, and vice versa. In practice this does not seem to work. There are a number of situations that we will encounter later (such as leading edge subspaces in Chapter 3 or approximate diagonalization in Chapter 6) where the authors have no idea of how to formulate a certain Weyr form argument in terms of the Jordan form.
- (2) With the benefit of the duality theorem and hindsight, the existence of the Weyr form seems fairly obvious if one already knows about the Jordan form.¹⁶ Being among a handful of authors who have rediscovered the Weyr form¹⁷ (see the historical remarks at the end of this section), we can assure you it is not (or at least not to us plodders) ! \square

Corollary 2.4.3

Let $\omega_1, \omega_2, \omega_3, \dots$ be the Weyr characteristic of a matrix $A \in M_n(F)$ associated with an eigenvalue λ , and let J be the Jordan form of A . Then for all positive integers k :

- (1) The number of basic Jordan blocks in J with eigenvalue λ and size at least $k \times k$ is ω_k .
- (2) The number of $k \times k$ basic Jordan blocks in J with eigenvalue λ is $\omega_k - \omega_{k+1}$.

Proof

(1) It is enough to establish this in the case A has a single eigenvalue λ . Let (n_1, n_2, \dots, n_r) and (m_1, m_2, \dots, m_s) be the Weyr and Jordan structures, respectively of A , which as we know are dual partitions of n . Recall that r is the nilpotent index of $A - \lambda I$, which in turn is m_1 , the size of the largest Jordan block in the Jordan form. Thus, the statement in (1) holds when $k > r$. Now assume $k \leq r$. Picture the Young tableau associated with (n_1, n_2, \dots, n_r) . The number of m_i that are at least k must be n_k , because m_i is the length of the i th column of the tableau and n_k is the length of the k th row. Finally, by Proposition 2.2.3, $\omega_k = n_k$.

- (2) Knowing (1), this should not be a challenge! \square

If we know a matrix in Weyr form of a nilpotent linear transformation $T: V \rightarrow V$ relative to some ordered basis \mathcal{B} , then the proof of Theorem 2.4.1 tells

16. This makes Weyr's original discovery even more remarkable, because it appears he did not know of the Jordan form.

17. But we did not observe the duality. That had to be pointed out to us! Thank you, Milen Yakimov.

us how to use dual Young diagrams to quickly reorder \mathcal{B} to obtain a basis \mathcal{B}' that gives the Jordan matrix for T . The converse works as well. Applied to a nilpotent matrix $A \in M_n(F)$ (by viewing A as a linear transformation of F^n under left multiplication), this procedure provides an explicit permutation matrix P that conjugates the Weyr form of A to its Jordan form (or vice versa). In particular, we have another proof for the existence of the Weyr form of A , if one accepts the existence of the Jordan form. Uniqueness of the Jordan form of A would also confirm uniqueness of the Weyr form. For if A had two Weyr forms W_1 and W_2 with different Weyr structures, the Jordan forms J_1, J_2 of W_1, W_2 would have different (dual) Jordan structures, contradicting the uniqueness of the Jordan form of A .¹⁸

Example 2.4.4

Suppose A is some 10×10 matrix having a single eigenvalue λ , and suppose we have computed its Weyr form W (for example, by the algorithm we shall present in the next section). That is, we have an explicit $C \in GL_{10}(F)$ with $W = C^{-1}AC$. Further suppose the Weyr structure turns out to be $(5, 3, 2)$. What is the quick way of getting the Jordan form J of A from this, via an explicit similarity transformation on A ? We simply need to calculate a certain permutation matrix P and let $J = P^{-1}WP = (CP)^{-1}A(CP)$.

Let $T: F^{10} \rightarrow F^{10}$ be a linear transformation whose matrix relative to some (ordered) basis $\mathcal{B} = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}\}$ is W . The proof of 2.4.1 tells us to do the following. First, form the Young diagram corresponding to the partition $(5, 3, 2)$ by distributing (in order) the basis vectors of \mathcal{B} from left to right across the rows of the Young tableau:

v_1	v_2	v_3	v_4	v_5
v_6	v_7	v_8		
v_9	v_{10}			

Next, form the dual Young diagram corresponding to the dual partition $(3, 3, 2, 1, 1)$, which by Theorem 2.4.1 is the Jordan structure of A :

v_1	v_6	v_9
v_2	v_7	v_{10}
v_3	v_8	
v_4		
v_5		

18. Compared to Section 2.2, these would be very roundabout arguments for the Weyr form if we had to start from scratch and first establish the existence and uniqueness of the Jordan form.

Now, form the ordered basis

$$\mathcal{B}' = \{v_1, v_6, v_9, v_2, v_7, v_{10}, v_3, v_8, v_4, v_5\} = \{v'_1, v'_2, \dots, v'_{10}\}$$

by running across the rows of the dual diagram (or down the columns of the original diagram if one prefers). Finally, let P be the permutation matrix corresponding to the reordering of the basis elements; that is, P is the matrix obtained by permuting the rows of the identity matrix according to the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 6 & 9 & 2 & 7 & 10 & 3 & 8 & 4 & 5 \end{pmatrix}.$$

Note that P is just the change of basis matrix $[\mathcal{B}', \mathcal{B}]$. Now $P^{-1}WP$ is in Jordan form because it is the matrix of T relative to \mathcal{B}' . Exactly the same procedure applies in going from the Jordan form to the Weyr form.

Rather than using dual Young diagrams, some may prefer to use a reordering of arrowed diagrams, as we did at the beginning of the chapter. Then we would start with the Weyr shifting on the v_i

$$\begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ v_1 & v_2 & v_3 & v_4 & v_5 \\ \uparrow & \uparrow & \uparrow & & \\ v_6 & v_7 & v_8 & & \\ \uparrow & \uparrow & & & \\ v_9 & v_{10} & & & \end{array}$$

and reorder the basis elements according to the column order to get \mathcal{B}' . The new Jordan arrow diagram (but expressed in terms of rows of \mathcal{B}' basis elements) is then:

$$\begin{array}{ccccccc} 0 & \leftarrow & v'_1 & \leftarrow & v'_2 & \leftarrow & v'_3 \\ 0 & \leftarrow & v'_4 & \leftarrow & v'_5 & \leftarrow & v'_6 \\ 0 & \leftarrow & v'_7 & \leftarrow & v'_8 & & \\ 0 & \leftarrow & v'_9 & & & & \\ 0 & \leftarrow & v'_{10} & & & & \end{array}$$

□

Remark 2.4.5

With the aid of Theorem 2.4.1, the reader may wish to check the answers for the Jordan structures of the “magnificent seven” matrices in Example 2.1.2. They are, respectively:

$$(5, 3, 1, 1)$$

$$(4, 4, 2)$$

$$(3, 3, 1)$$

$$(2, 2, 2)$$

$$(3, 3)$$

$$(4)$$

$$(1, 1, 1, 1)$$

□

As a corollary to Theorem 2.4.1, we can now show, fairly painlessly, how the Jordan structure of a Jordan matrix J with a single eigenvalue λ is completely determined by the nullities of the powers of $J - \lambda I$, by using the corresponding result for the Weyr form (Proposition 2.2.3).

Corollary 2.4.6

Let J be a Jordan matrix with a single eigenvalue λ and with Jordan structure (m_1, m_2, \dots, m_s) . Then:

$$s = \text{nullity}(J - \lambda I),$$

$$m_1 = \text{nilpotency index of } J - \lambda I,$$

$$m_i = \text{number of integers } j \text{ between 1 and } s \text{ such that}$$

$$\text{nullity}(J - \lambda I)^j - \text{nullity}(J - \lambda I)^{j-1} \geq i,$$

$$\text{for } i = 2, 3, \dots, s.$$

Proof

On the one hand, by Proposition 2.2.3 we know that the Weyr structure of J is (n_1, n_2, \dots, n_r) where r is the nilpotent index of $J - \lambda I$ and $n_j = \text{nullity}(J - \lambda I)^j - \text{nullity}(J - \lambda I)^{j-1}$ for $j = 1, 2, \dots, r$. On the other hand, Theorem 2.4.1 tells us that (m_1, m_2, \dots, m_s) is the dual partition of (n_1, n_2, \dots, n_r) . But by the very nature of a dual partition, $s = n_1$, $m_1 = r$, and, for $i = 1, 2, \dots, s$

$$m_i = \# \text{ integers } j \text{ between 1 and } s \text{ such that } n_j \geq i.$$

□

It is not often that the Weyr and Jordan forms of a matrix are the same.¹⁹ The following proposition characterizes the situation. We leave its proof as an exercise (we will not be using the result).

Proposition 2.4.7

- (1) *A matrix J that is in Jordan form is also in Weyr form if and only if for each eigenvalue λ of J , there is just one basic λ -block or all its basic λ -blocks are 1×1 .*
- (2) *Over an algebraically closed field, the Jordan and Weyr forms of a square matrix A agree (as unblocked matrices) if and only if to within similarity, A is a direct sum of a nonderogatory matrix B and a diagonalizable matrix C that do not share a common eigenvalue.*

Historical Remarks

Now that we know the connection of the Weyr form to the Jordan form, we are in a position to make some historical remarks on how the Weyr form and its competing terms have evolved. (Weyr himself never assigned a name to his form, let alone his own name, and referred to canonical matrix forms simply as “typical matrices.”) We thank Roger Horn for pointing out some of this history. As we have earlier commented, the Weyr form has been rediscovered several times since Weyr’s original discovery in 1885, with those later investigators driven principally by the desire for the centralizing matrices to be block upper triangular.

Belitskii rediscovered the Weyr form in his 1983 paper (English version, 2000) but he used the term “modified Jordan matrix.” He observed that his form was permutationally similar to the Jordan form (our Theorem 2.4.1), and he was probably the first to observe the nice block upper triangular form of the centralizing matrices (our Proposition 2.3.3). The latter makes the centralizer $\Gamma = \mathcal{C}(W)$ of an $n \times n$ Weyr matrix W a “reduced algebra,” meaning Γ is a subalgebra of $M_n(F)$ whose members take a certain block upper triangular form. (Here, F is any algebraically closed field.) The 1983 paper contains the important “Belitskii algorithm” for establishing a Γ -canonical form M^∞ for a matrix $M \in M_n(F)$ relative to a given reduced algebra $\Gamma \subseteq M_n(F)$. Two $n \times n$ matrices M and N are Γ -similar (similar under conjugation by some invertible member of Γ) if and only if $M^\infty = N^\infty$. This algorithm has many applications.

19. However, even when the two forms are different, they share an interesting property within their similarity class. Each is optimal with respect to the number of nonzero off-diagonal entries. This is known for the Jordan form (see footnote 26 on p.40) and therefore must also be true for the Weyr form because, being permutationally similar to the Jordan form, it has the same number of nonzero entries and the same diagonal entries (in some order).

For instance, applied to $\Gamma = \mathcal{C}(W)$, where W is a prescribed Weyr matrix, the algorithm produces canonical pairs (W, C) of matrices with respect to (full) similarity from the set of all pairs (W, B) of (not necessarily commuting) $n \times n$ matrices where the first is the prescribed Weyr matrix. Doesn't that invite interesting applications? Belitskii also established that *every* subalgebra of $M_n(F)$ is similar to some reduced algebra.

The first popular account of the Weyr characteristic and Weyr form was in 1999, when Shapiro wrote an article "The Weyr Characteristic" for the American Mathematical Monthly.²⁰ There she described the Weyr canonical form under that very name.²¹ So historically, perhaps Shapiro gets the credit for finally nailing the correct term for this particular canonical form.²² Interestingly, Shapiro says she first learned of the Weyr characteristic in 1980 from Hans Schneider,²³ who had used it for studying the singular graph of an M-matrix. Shapiro mentions the importance of the Weyr characteristic in computing the Jordan form of a complex matrix in a stable manner, and how to obtain the Weyr characteristic using unitary similarity (see our Remark 2.5.2 in the following section). However, perhaps due to space restrictions, Shapiro does not mention or hint at any applications of the Weyr form itself, so some readers may have been left wondering "why bother with this form." In particular, there is no mention of the characterization of matrices that centralize a nilpotent Weyr matrix, the most important property for our later applications in Part II.

Sergeichuk extended Belitskii's algorithm in 2000 and applied it to a broad class of important matrix problems, related to classifying representations of quivers, posets, and finite-dimensional algebras. It is an impressive piece of work. Also, this paper (more accurately its preprint) may have been the first to use the term "Weyr matrix" and stress its connection with the Weyr characteristic. In his earlier work, Sergeichuk had used terms such as "reordered Jordan matrix" or "modified Jordan matrix."

20. This publication has a very large readership across a broad spectrum of people interested in mathematics generally, and so it presents an ideal forum for promoting concepts with widespread applications.

21. It is a pity that the title of her article did not also draw attention to a matrix canonical form that is related to the Jordan form. That may well have helped others to come to know the Weyr form.

22. Prior to writing her article, Shapiro was in receipt of a preprint of Sergeichuk's 2000 paper, so she may have been influenced by Sergeichuk's use of the term "Weyr matrix."

23. To round out the story of the "University of Otago connection" in footnote 24 on p. 37, we mention that the distinguished linear algebraist Hans Schneider was a Ph.D student of A. C. Aitken at the University of Edinburgh. His degree was completed in 1952.

O'Meara and Vinsonhaler rediscovered the Weyr form in 2006 under the name “H-form,”²⁴ and also rediscovered many of its known properties. In that paper they also discovered the simultaneous upper triangularization of commuting matrices that puts the first matrix in Weyr form (Theorem 2.3.5). (It would appear that this was not previously known.)

Harima and Watanabe rediscovered the Weyr form in 2008 under the name “Jordan second canonical form” in a study related to commutative Artinian algebras. They too were driven by the form of the centralizer of a canonical matrix. In their paper, they say the idea for their canonical form was suggested to them by Weyl’s 1946 book *Classical Groups, Their Invariants and Representations*. □

2.5 COMPUTING THE WEYR FORM

In some applications of the Weyr form, such as in Chapter 5, just the existence of the Weyr form suffices. In other applications, however, such as in Section 3.5 of Chapter 3, and some applications in Chapter 6, we need to actually find a similarity transformation that puts a given matrix $A \in M_n(F)$ in Weyr form, that is, an explicit $C \in GL_n(F)$ such that $C^{-1}AC$ is in Weyr form. In this section we present an algorithm for doing this, as well as examples to illustrate the relatively straightforward individual calculations involved (easier than for the Jordan form). To see that the algorithm works, one can look back at the proof of Theorem 2.2.2.

Fix an algebraically closed field F and an $n \times n$ matrix A over F . The first step in computing the Weyr form of A is no different from that for the Jordan form—one needs to know the eigenvalues of A . There is no way around that. Of course, computing eigenvalues is a subject in its own right, to which we will add nothing. We will just assume we know the distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$. Then, as with the Jordan form, we use the Generalized Eigenspace Decomposition 1.5.2 (more specifically, its Corollary 1.5.4) to reduce to the case where A has only a single eigenvalue λ (this involves calculating bases for various $\ker(\lambda_i I - A)^{m_i}$). By subtraction of the scalar matrix λI , we are down to the nilpotent case. Our algorithm applies to the nilpotent case.

Algorithm for computing the Weyr form of a nilpotent matrix A

Step 1. Set $A_1 = A$. So far, so good.²⁵

24. “H” stood for “Husky” in recognition of the University of Connecticut connection.

25. It is hard to argue that this is not a reasonable step !

Step 2. If A_1 is nonzero, compute a basis for the null space of A_1 (by elementary row operations) and extend it to a basis for F^n , where n is the matrix size of A_1 . Let P_1 be the $n \times n$ matrix having the latter basis vectors as its columns. Then

$$P_1^{-1}A_1P_1 = \left[\begin{array}{c|c} 0 & B_2 \\ \hline 0 & A_2 \end{array} \right],$$

where A_2 is a square matrix of size $n - \text{nullity } A_1$.

Step(s) 3. If A_2 is nonzero, repeat Step 2 on A_2 to obtain an invertible matrix P_2 of size $n - \text{nullity } A_1$ such that

$$P_2^{-1}A_2P_2 = \left[\begin{array}{c|c} 0 & B_3 \\ \hline 0 & A_3 \end{array} \right],$$

where A_3 is a square matrix of size $n - \text{nullity } A_1 - \text{nullity } A_2$. Continue this process of producing decreasingly smaller square matrices A_1, A_2, \dots and associated invertible matrices P_1, P_2, \dots (of matching size) until the first zero matrix A_r results. (This must happen since A is nilpotent.) Then the Weyr structure of A is (n_1, n_2, \dots, n_r) , where $n_i = \text{nullity } A_i$ for $i = 1, 2, \dots, r$.

Step 4. Conjugate A by the product of the invertible $n \times n$ matrices

$$P_1, \text{diag}(I, P_2), \text{diag}(I, P_3), \dots, \text{diag}(I, P_{r-1})$$

(for appropriately sized identity matrices I). This results in an $r \times r$ blocked matrix

$$X = \left[\begin{array}{cccccc} 0 & X_{12} & X_{13} & X_{14} & \cdots & X_{1r} \\ & 0 & X_{23} & X_{24} & \cdots & \vdots \\ & & 0 & X_{34} & \cdots & \vdots \\ & & & & \ddots & \vdots \\ & & & & & 0 & X_{r-1,r} \\ & & & & & & 0 \end{array} \right],$$

where X_{ij} is an $n_i \times n_j$ matrix (whose diagonal blocks of zeros are $n_1 \times n_1, \dots, n_r \times n_r$). The first superdiagonal blocks have full column-rank. (That is, $\text{rank } X_{i,i+1} = n_{i+1}$ for $i = 1, 2, \dots, r-1$.)

Step 5. Using elementary row operations, compute $Y_{r-1} \in GL_{n_{r-1}}(F)$ such that

$$Y_{r-1}X_{r-1,r} = \begin{bmatrix} I \\ 0 \end{bmatrix},$$

where I is the $n_r \times n_r$ identity. Conjugate X by $Q_1 = \text{diag}(I, I, \dots, Y_{r-1}^{-1}, I)$ to convert $X_{r-1,r}$ to

$$I_{r-1,r} = \begin{bmatrix} I \\ 0 \end{bmatrix},$$

the $n_{r-1} \times n_r$ matrix having the $n_r \times n_r$ identity matrix as its upper part and zero rows below. This preserves the form of X (the only other blocks changed are in column $r-1$). Use conjugations by a product R_1 of elementary matrices to clear out all the blocks above $I_{r-1,r}$ in the last column of blocks (see Lemma 2.2.1). This preserves the form of X (changing only the r th column). It is possible to write down R_1 explicitly (essentially as a product of “elementary block matrices”):

$$R_1 = I + \begin{bmatrix} 0 & \cdots & 0 & \overline{X}_{1r} & 0 \\ 0 & \cdots & 0 & \overline{X}_{2r} & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \overline{X}_{r-2,r} & 0 \\ 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 \end{bmatrix},$$

where \overline{X}_{ir} is X_{ir} with $n_{r-1} - n_r$ zero columns appended.

Step(s) 6. Repeat Step 5 on column $r-1$, converting the $(r-2, r-1)$ block to $I_{r-2,r-1}$ via conjugation by some $Q_2 \in GL_n(F)$. Use this “identity” block to clear out the blocks above, via conjugation by a product R_2 of elementary matrices. This doesn’t change the last column of blocks and preserves the form of X . Repeat this process on columns $r-2, r-3, \dots, 3, 2$, using conjugations

by $Q_3, R_3, \dots, Q_{r-2}, R_{r-2}, Q_{r-1}$. The resulting matrix W is now in Weyr form. Let

$$C = P_1 \operatorname{diag}(I, P_2) \cdots \operatorname{diag}(I, P_{r-1}) Q_1 R_1 Q_2 \cdots R_{r-2} Q_{r-1}.$$

Then $W = C^{-1}AC$ gives the desired explicit similarity transformation. As in Step 5, we can also write down R_2, R_3, \dots explicitly. \square

Remark 2.5.1

In practice, it is not necessary to stick slavishly to Steps 1–6. Shortcuts will appear, depending on the particular matrix A . There are just two phases to remember:

- (1) Transform A to a strictly block upper triangular matrix $X = (X_{ij})$ in which the diagonal blocks are decreasing in size, and the first superdiagonal blocks $X_{12}, X_{23}, \dots, X_{r-1,r}$ have full column-rank.
- (2) Starting with the *last* column of blocks and working backwards, convert the first superdiagonal block to an “identity” matrix, and use that to clear out the blocks above it.

If all one is interested in is the Weyr structure of A , that is presented in phase (1) by the diagonal block sizes. \square

Remark 2.5.2

Two complex matrices $A, B \in M_n(\mathbb{C})$ are called *unitarily similar* if $B = U^{-1}AU$ for some unitary matrix U . Recall that the latter is an invertible matrix U whose inverse is the conjugate transpose U^* (equivalently, the columns of U are orthonormal vectors). A unitary transformation $X \mapsto U^*XU$ of $M_n(\mathbb{C})$ not only preserves algebraic properties but also “geometric” ones. When $F = \mathbb{C}$, the first phase of our algorithm for computing the Weyr form of a nilpotent matrix can be achieved via a unitary transformation (by extending orthonormal bases for null spaces to orthonormal bases for the full space). This is very useful for numerical accuracy, because there are then no roundoff errors introduced in the inverse (= conjugate transpose) of the unitary matrix. On the other hand, in general the second phase cannot be accomplished by a unitary transformation, because an $n \times n$ complex matrix $A = (a_{ij})$ is not usually unitarily similar to its Weyr form (or Jordan form). A quick way of seeing this is in terms of a matrix norm $\|A\|$ defined by

$$\|A\|^2 = \sum_{i,j=1}^n |a_{ij}|^2.$$

(We will meet this norm again in Chapter 6.) Since $\|A\|^2 = \text{tr}(AA^*)$ and a unitary transformation preserves traces, products, and conjugate transposes, a pair of unitarily similar matrices must have the same norm. In particular, the nilpotent matrix

$$A = \left[\begin{array}{cc|cc} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & 1 \\ & & & 0 \end{array} \right],$$

whose Weyr structure is $(2, 1, 1)$, can't be unitarily similar to its Weyr form W because $\|A\| = \sqrt{3}$ whereas $\|W\| = \sqrt{2}$. (Note that A has the form required for the second phase of our algorithm.) \square

To illustrate the algorithm, we look at a couple of examples. The first is fairly simple.

Example 2.5.3

We wish to put the 4×4 nilpotent matrix

$$A = \begin{bmatrix} 2 & -1 & 3 & 1 \\ 1 & -1 & 2 & 1 \\ -2 & 1 & -3 & -1 \\ 3 & -2 & 5 & 2 \end{bmatrix}$$

in Weyr form. (As we'll see, $A^2 = 0$.) Set $A_1 = A$. By elementary row operations, we have

$$\begin{aligned} A_1 &\longrightarrow \begin{bmatrix} 1 & -1 & 2 & 1 \\ 2 & -1 & 3 & 1 \\ -2 & 1 & -3 & -1 \\ 3 & -2 & 5 & 2 \end{bmatrix} \\ &\longrightarrow \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & -1 & 1 & 1 \\ 0 & 1 & -1 & -1 \end{bmatrix} \\ &\longrightarrow \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

From this, we see that nullity $A_1 = 2$ and that the set

$$\left\{ \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for F^4 in which the first two vectors form a basis for the null space of A_1 . (It pays to extend a basis as simply as possible, by throwing in standard basis vectors.) Putting these basis vectors as columns we get the invertible matrix

$$P_1 = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

for which

$$P_1^{-1}A_1P_1 = \left[\begin{array}{cc|cc} 0 & 0 & -3 & -1 \\ 0 & 0 & 5 & 2 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right].$$

(The quick way to get this is to calculate the matrix of A_1 under a change of basis—we know the first two columns are zero, and the last two come from expressing the 3rd and 4th columns of A_1 in terms of the new basis.)

Since the bottom right-hand block A_2 of the transformed A_1 is zero, we can deduce that the Weyr structure of A is $(2, 2)$ and so, at this stage, we can write down the Weyr form of A . In terms of our algorithm, Steps 3 and 4 are not required. Thus, we move to Step 5. (Step 6 will not be needed.) Let X_{12} be the $(1, 2)$ block of $X = P_1^{-1}A_1P_1$ (where the label X conforms to the algorithm). As predicted by the algorithm, X_{12} has rank 2 and we can convert X_{12} to I using conjugation by

$$Q_1 = \text{diag}(X_{12}, I) = \left[\begin{array}{cc|cc} -3 & -1 & 0 & 0 \\ 5 & 2 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right].$$

Thus, conjugating A by the invertible matrix

$$C = P_1 Q_1 = \begin{bmatrix} 3 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -3 & -1 & 1 & 0 \\ 5 & 2 & 0 & 1 \end{bmatrix}$$

produces the desired Weyr form W of A :

$$W = C^{-1}AC = \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right].$$

□

Example 2.5.4

Our second example is a little more involved yet can be done comfortably using hand calculations.²⁶ Of course, Maple or Matlab can be used if one prefers. We wish to put the 7×7 nilpotent matrix

$$A = \begin{bmatrix} -1 & 0 & 0 & 2 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 6 & -2 & -2 & 4 & 2 & -2 & 4 \\ -2 & 1 & 1 & 0 & 0 & 1 & -1 \\ 3 & -1 & -1 & 2 & 1 & -1 & 2 \\ -5 & 2 & 2 & -5 & -3 & 2 & -4 \\ 2 & 0 & 0 & -4 & -2 & 0 & 0 \end{bmatrix}$$

in Weyr form, following the steps in our algorithm. (We'll see that A has nilpotency index 3.) Set $A_1 = A$, the first step.²⁷ By elementary row operations

$$A_1 \longrightarrow \begin{bmatrix} 1 & 0 & 0 & -2 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -2 & -2 & 16 & 8 & -2 & 4 \\ 0 & 1 & 1 & -4 & -2 & 1 & -1 \\ 0 & -1 & -1 & 8 & 4 & -1 & 2 \\ 0 & 2 & 2 & -15 & -8 & 2 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

26. Besides, to paraphrase Frankie Laine's "Moonlight Gambler," if one has never hand calculated a canonical form, one has never calculated at all !

27. Maple and Matlab should not be required for this step.

$$\longrightarrow \begin{bmatrix} 1 & 0 & 0 & -2 & -1 & 0 & 0 \\ 0 & 1 & 1 & -4 & -2 & 1 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 4 & 0 & 2 \\ 0 & 0 & 0 & 4 & 2 & 0 & 1 \\ 0 & 0 & 0 & -7 & -4 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 1 & 0 & 0 & -2 & -1 & 0 & 0 \\ 0 & 1 & 1 & -4 & -2 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 2 \\ 0 & 0 & 0 & 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & -4 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 1 & 0 & 0 & -2 & -1 & 0 & 0 \\ 0 & 1 & 1 & -4 & -2 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

From this, we see that the nullity of A_1 is $n_1 = 3$ and that the set

$$\left\{ \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

is a basis for F^7 in which the first three vectors form a basis for the null space of A_1 . (We have chosen to extend the basis by throwing in the 1st, 3rd, 4th, and 5th standard basis vectors.) Let $P_1 \in GL_7(F)$ be the matrix having the above basis vectors as its columns. By change of basis calculations (or direct evaluation in

Matlab or Maple), we find

$$P_1^{-1}A_1P_1 = \left[\begin{array}{ccc|cccc} 0 & 0 & 0 & 6 & -2 & 4 & 2 \\ 0 & 0 & 0 & -5 & 2 & -5 & -3 \\ 0 & 0 & 0 & 1 & 0 & -2 & -1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 4 & -1 & 0 & 0 \end{array} \right].$$

Let A_2 be the bottom right-hand block:

$$A_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 4 & -1 & 0 & 0 \end{bmatrix}$$

This completes Step 2 of the algorithm. Since A_2 is nonzero, we move on to Step 3.

Step 3 says to repeat Step 2 on A_2 . However, as remarked in 2.5.1, one should be on the lookout for shortcuts depending on the actual matrix. There is one staring us in the face. Clearly, the nullity of A_2 is $n_2 = 2$, and we can obtain the desired form for Step 3 just by conjugating A_2 by

$$P_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

(This will swap the two off-diagonal 2×2 blocks, if we view A_2 as a blocked matrix.) Thus,

$$P_2^{-1}A_2P_2 = \left[\begin{array}{cc|cc} 0 & 0 & -2 & 1 \\ 0 & 0 & 4 & -1 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right].$$

The bottom right-hand block A_3 is now the zero 2×2 matrix. We set $n_3 = 2$ (the nullity of A_3) and move directly to Step 4. Note that we now know the Weyr structure of A is $(n_1, n_2, n_3) = (3, 2, 2)$.

Conjugating $P_1^{-1}A_1P_1$ by $\text{diag}(I, P_2)$ gives the blocked matrix

$$X = \left[\begin{array}{ccc|cc|cc} 0 & 0 & 0 & 4 & 2 & 6 & -2 \\ 0 & 0 & 0 & -5 & -3 & -5 & 2 \\ 0 & 0 & 0 & -2 & -1 & 1 & 0 \\ \hline & & & 0 & 0 & -2 & 1 \\ & & & 0 & 0 & 4 & -1 \\ \hline & & & & & 0 & 0 \\ & & & & & 0 & 0 \end{array} \right]$$

in which the first superdiagonal blocks have full column-rank. This completes Step 4. We are ready for the second phase of the reduction, converting the first superdiagonal blocks to “identity” matrices and clearing out the blocks above.

To convert the (invertible) $(2, 3)$ block

$$X_{23} = \begin{bmatrix} -2 & 1 \\ 4 & -1 \end{bmatrix}$$

to I , we conjugate X by $Q_1 = \text{diag}(I, X_{23}, I)$ to obtain

$$Q_1^{-1}XQ_1 = \left[\begin{array}{ccc|cc|cc} 0 & 0 & 0 & 0 & 2 & 6 & -2 \\ 0 & 0 & 0 & -2 & -2 & -5 & 2 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ \hline & & & 0 & 0 & 1 & 0 \\ & & & 0 & 0 & 0 & 1 \\ \hline & & & & & 0 & 0 \\ & & & & & 0 & 0 \end{array} \right].$$

By elementary row operations, we can clear out the entries above the converted block. Specifically, left multiplying by the matrix

$$E_{34}(-1)E_{24}(5)E_{14}(-6)E_{25}(-2)E_{15}(2)$$

will do this, where $E_{ij}(c)$ denotes the elementary matrix $I + ce_{ij}$ for $i \neq j$. But by Lemma 2.2.1, this multiplication is the same as conjugating by the inverse of

the product,²⁸ namely by

$$R_1 = E_{15}(-2)E_{25}(2)E_{14}(6)E_{24}(-5)E_{34}(1)$$

$$= \begin{bmatrix} 1 & 0 & 0 & 6 & -2 & 0 & 0 \\ 0 & 1 & 0 & -5 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Thus,

$$R_1^{-1}Q_1^{-1}XQ_1R_1 = \left[\begin{array}{ccc|cc|cc} 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & -2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ \hline & & & 0 & 0 & 1 & 0 \\ & & & 0 & 0 & 0 & 1 \\ \hline & & & & & 0 & 0 \\ & & & & & 0 & 0 \end{array} \right].$$

This completes Step 5.

Our last Step 6 is to fix up the $(1, 2)$ block of the above matrix. Putting the 3×2 block in reduced row-echelon form can be achieved by left multiplying by a 3×3 invertible matrix:

$$\frac{1}{2} \begin{bmatrix} 0 & -1 & 2 \\ 0 & 0 & -2 \\ 2 & 0 & 4 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ -2 & -2 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

In terms of the full 7×7 matrix, we can achieve this change on the $(1, 2)$ block by conjugating with

$$Q_2 = \text{diag}(Y, I)$$

28. Alternatively, we can produce R_1 directly as a product of “elementary block matrices,” as described in Step 5 of the algorithm.

where

$$Y = 2 \begin{bmatrix} 0 & -1 & 2 \\ 0 & 0 & -2 \\ 2 & 0 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 2 & 1 \\ -2 & -2 & 0 \\ 0 & -1 & 0 \end{bmatrix}.$$

Finally, we have our Weyr form W for A :

$$W = C^{-1}AC = \left[\begin{array}{ccc|cc|cc} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline & & & 0 & 0 & 1 & 0 \\ & & & 0 & 0 & 0 & 1 \\ \hline & & & & & 0 & 0 \\ & & & & & 0 & 0 \end{array} \right]$$

where

$$C = P_1 \operatorname{diag}(I, P_2) Q_1 R_1 Q_2$$

$$= \begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 1 & 0 \\ 2 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 6 & -2 & 0 & 1 \\ 0 & 0 & 0 & -2 & 1 & 0 & 0 \\ 0 & 1 & 0 & 3 & -1 & 0 & 0 \\ -2 & -2 & 0 & -5 & 2 & 0 & 0 \\ 0 & -2 & 0 & 2 & 0 & 0 & 0 \end{bmatrix}.$$

By duality, the Jordan structure of A is $(3, 3, 1)$. It is a simple task to follow up our similarity transformation with an explicit permutation transformation that puts A in Jordan form (see discussion in Example 2.4.4). We leave that as an exercise. \square

We close this chapter with our second test question, designed to check one's understanding of the first phase of the algorithm. The numerical calculations are very straightforward. A successful outcome should entice the reader to proceed quickly to Chapter 3.

Test Question 2. What is the Weyr form W of the following 6×6 matrix A whose only eigenvalue is -3 ? The answer is given below. (A keen reader

may wish to use the second phase of the algorithm to actually find a similarity transformation that converts A to W .)

$$A = \begin{bmatrix} -3 & 1 & 0 & 0 & -3 & 3 \\ 0 & -7 & 0 & 0 & 6 & -10 \\ 0 & 8 & -3 & 0 & -10 & 17 \\ 0 & -3 & 0 & -3 & 5 & -7 \\ 0 & -6 & 0 & 0 & 6 & -15 \\ 0 & -2 & 0 & 0 & 3 & -8 \end{bmatrix}$$

ANSWER TO TEST QUESTION 1 (after Remark 2.1.7). Only the second matrix is in Weyr form. The first matrix is the direct sum of two basic Weyr matrices but with the same eigenvalue. The third matrix has a “slipped identity disc” in its $(1, 2)$ block.

ANSWER TO TEST QUESTION 2.

$$W = \left[\begin{array}{ccc|cc|c} -3 & 0 & 0 & 1 & 0 & 0 \\ 0 & -3 & 0 & 0 & 1 & 0 \\ 0 & 0 & -3 & 0 & 0 & 0 \\ \hline & & & -3 & 0 & 1 \\ & & & 0 & -3 & 0 \\ & & & & & -3 \end{array} \right]$$

BIOGRAPHICAL NOTE ON WEYR

Eduard Weyr was born on June 22, 1852, in Prague (in Bohemia, now the Czech Republic). His father was a mathematician at a secondary school in Prague. (His older brother Emil was also to become a great mathematician.) Eduard studied at his father’s school before attending the Prague Polytechnic and Charles-Ferdinand University, also in Prague. He had already sent two papers to the academy in Vienna by the time he was 16. Following his Prague studies he traveled to Göttingen, receiving his doctorate there in 1873 with a thesis titled *Über algebraische Raumkurven*. After a short spell in Paris studying under Hermite and Serret, he returned to Prague where he eventually became a professor at Charles-Ferdinand University. The Weyr form appears briefly

in his 1885 *Comptes Rendus* paper “Répartition des matrices en espèces et formation de toutes les espèces” and in more detail in the much-longer “Zur Theorie der bilinearen Formen,” in *Monatsh. Math. Physik* in 1890. The latter paper is a wonderful piece of mathematics for its time, modern and clear even by today’s standards. It is arguably the first paper in linear algebra, as distinct from matrix theory. It is interesting that Weyr cites the work of Frobenius, Sylvester, Cauchy, and Hermite in canonical forms but never mentions Jordan in this context! However, in his work on the history of mathematics, Brechenmacher notes that in the period 1885–1890 Weyr was the only mathematician on the European continent using Cayley and Sylvester’s pioneering work on matrices. So, was Weyr aware of the Jordan form? Our guess is probably not initially in the matrix setting—Jordan’s result appeared in the (nineteenth-century) language of permutation group theory and did not evolve into the canonical matrix form of choice until the 1930s. In the meantime, Weyr’s form sank into obscurity. It would appear that Weyr himself never really appreciated the utility of his own form in commutativity problems, such as we study in Part II of our book. Weyr also published research in geometry, in particular projective and differential geometry. He died in Zábory, Bohemia, on July 23, 1903.

Centralizers

The Weyr form is superior to the Jordan form when it comes to centralizers. In this chapter we lay the groundwork for our bold contention. But it is the applications in Part II of our book that will finally substantiate the claim.

For a given $n \times n$ matrix A over F , its **centralizer** $\mathcal{C}(A)$ is the subalgebra of $M_n(F)$ consisting of all matrices B that commute with A . The study of the centralizer quickly reduces to the nilpotent case. In this chapter, we study in more detail the precise form of the centralizer of a nilpotent matrix A , depending on whether A is in Jordan form or Weyr form. In particular, in Sections 3.1 and 3.2, we compute the dimension of $\mathcal{C}(A)$ in each case. For the Jordan form, this is known as the *Frobenius formula*. The corresponding formula for the Weyr form involves a sum of squares. Equating the two gives a matrix structure insight into an interesting number-theoretic identity used by Gerstenhaber, as is discussed in Section 3.3. That aside, however, much of the material developed in this chapter is for use in Chapters 5, 6, and 7. **Familiarity with the ideas in the present chapter is essential for a full understanding of these later chapters.**

Throughout this chapter, F denotes an algebraically closed field.

In Section 3.4, we show that if W is a nilpotent matrix in Weyr form and with Weyr structure (n_1, n_2, \dots, n_r) , then we can associate with each subalgebra \mathcal{A} of $\mathcal{C}(W)$, certain natural “leading edge” subspaces U_i of the space of $n_1 \times n_{i+1}$

matrices for $i = 0, 1, \dots, r-1$ such that $\dim \mathcal{A} = \dim U_0 + \dim U_1 + \dots + \dim U_{r-1}$. This turns out to be a most useful formula. A particular application of the formula is in computing the dimension of a commutative subalgebra \mathcal{A} of $M_n(F)$ containing a nilpotent Weyr matrix W of known Weyr structure. We will see applications of leading edge subspaces in Chapters 5, 6, and 7. To ease the reader into these, we look in some detail at several numerical examples in Section 3.5 of the present chapter. The reader who is feeling comfortable with the material up to this point can just skim, or even skip, the examples. As far as we know, our treatment of leading edge subspaces has not previously appeared in the literature.

Of course, because of duality, there are corresponding spaces of matrices that could be associated with a subalgebra of $\mathcal{C}(J)$ when J is a nilpotent matrix in Jordan form. However, these spaces are far less natural and difficult to even remember. For this reason, we shall not treat the Jordan analogue. It is one of a number of situations in the book where the Weyr form clearly trumps its Jordan counterpart.

3.1 THE CENTRALIZER OF A JORDAN MATRIX

We begin by recording why the study of centralizers reduces to the nilpotent case. It's all because of the Corollary 1.5.4 to the Generalized Eigenspace Decomposition 1.5.2 (together with the simple observation that similar matrices have isomorphic centralizers).

Proposition 3.1.1

Let F be an algebraically closed field. Suppose $A \in M_n(F)$ is a block diagonal matrix $\text{diag}(A_1, \dots, A_k)$, where each $A_i \in M_{m_i}(F)$ has a single eigenvalue λ_i with $\lambda_i \neq \lambda_j$ when $i \neq j$. Then the matrices that centralize A are precisely those of the form $B = \text{diag}(B_1, B_2, \dots, B_k)$, where each $B_i \in M_{m_i}(F)$ centralizes A_i . Consequently, as algebras

$$\mathcal{C}(A) \cong \prod_{i=1}^k \mathcal{C}(A_i).$$

Also, the centralizer of A_i (within the algebra $M_{m_i}(F)$) is the same as the centralizer of the nilpotent matrix $A_i - \lambda_i I$.

Proof

Suppose $B = \text{diag}(B_1, B_2, \dots, B_k)$ is a block diagonal matrix with $m_i \times m_i$ diagonal blocks. If each B_i centralizes A_i , then clearly B centralizes A .

Conversely, suppose $B \in M_n(F)$ centralizes A . Write $B = (B_{ij})$ as a $k \times k$ block matrix with the same block structure as A . Fix indices i and j with $i \neq j$. From

commutativity of A and B we have

$$A_i B_{ij} = B_{ij} A_j$$

whence $B_{ij} = 0$ by Sylvester's Theorem 1.6.1, because A_i and A_j have no common eigenvalues. Therefore, B is block diagonal. Noting the manner in which block diagonal matrices multiply, we now see that the i th diagonal block of B must centralize that of A .

The final two statements of the proposition should be clear. (Note that scalar matrices commute with everything.)¹ \square

Here now is the description of the matrices that centralize a nilpotent matrix in Jordan form.

Proposition 3.1.2

Let J be an $n \times n$ nilpotent matrix in Jordan form and with Jordan structure (m_1, m_2, \dots, m_s) . Let K be an $n \times n$ matrix, blocked according to $m_i \times m_i$ diagonal blocks, and let K_{ij} denote its (i, j) block for $i, j = 1, \dots, s$. Then J and K commute if and only if each of the s^2 blocks K_{ij} is upper triangular with northwest-southeast main diagonal and superdiagonals constant. That is, for $i \geq j$ each K_{ij} is of the form

$$K_{ij} = \begin{bmatrix} 0 & \cdots & 0 & a & \cdots & x & y & z \\ 0 & & & a & \cdots & x & y & \\ 0 & & & & \ddots & & x & \\ \vdots & & & & & & \vdots & \\ 0 & \cdots & 0 & & \cdots & a & & \end{bmatrix}$$

(with the first $m_j - m_i$ columns zero), and for $i \leq j$

$$K_{ij} = \begin{bmatrix} a & b & c & \cdots & \cdots & z \\ 0 & a & b & c & \cdots & \\ 0 & 0 & a & b & & \\ \vdots & & & \ddots & & \\ 0 & 0 & 0 & \cdots & a & \\ 0 & 0 & 0 & \cdots & 0 & \\ \vdots & \vdots & \vdots & & \vdots & \\ 0 & 0 & 0 & \cdots & 0 & \end{bmatrix}$$

1. The converse also holds: if $X \in M_n(F)$ commutes with all $Y \in M_n(F)$, then $X = \lambda I$ for some scalar λ .

(with the last $m_i - m_j$ rows zero). To ease notation, we have not indicated the dependence of the entries a, b, c, \dots, z on i, j .

Proof

As a block diagonal matrix, $J = \text{diag}(J_1, J_2, \dots, J_s)$ where J_i is the basic $m_i \times m_i$ Jordan matrix

$$J_i = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 0 & 1 \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix}.$$

The condition that K commutes with J is easily seen to be that

$$(*) \quad J_i K_{ij} = K_{ij} J_j \text{ for } i, j = 1, \dots, s.$$

But now recall the upward shifting effect on the rows of a matrix under left multiplication by J_i , and the right shifting effect on the columns of a matrix under right multiplication by J_j (see Remark 2.3.1). Thus, $(*)$ implies that all the northwest-southeast diagonals (upper and lower) of K_{ij} must be constant. But in the case $i \leq j$, it also says that the first column entries of K_{ij} , apart from the $(1, 1)$ entry, must be zero, whence in fact all the lower diagonals are zero. And in the case $i \geq j$, the $(*)$ condition says the last row entries of K_{ij} , apart from the (m_i, m_j) entry, must be zero so again all the lower diagonal entries are zero. In each case, K_{ij} is upper triangular with constant diagonal and superdiagonals. Conversely, it is straightforward to confirm that this condition implies $(*)$. \square

We next derive the Frobenius formula for the dimension of the centralizer of a nilpotent Jordan matrix.

Proposition 3.1.3 (Frobenius Formula)

Let A be a nilpotent $n \times n$ matrix over a field F , and let (m_1, m_2, \dots, m_s) be the Jordan structure of A . Then

$$\dim \mathcal{C}(A) = m_1 + 3m_2 + 5m_3 + \dots + (2s - 1)m_s.$$

Proof

We can assume A is already in Jordan form because a similarity transformation won't alter the dimension of the centralizer. (If $B = C^{-1}AC$, then $\mathcal{C}(B) = C^{-1}\mathcal{C}(A)C \cong \mathcal{C}(A)$, whence $\dim \mathcal{C}(B) = \dim \mathcal{C}(A)$.) By Proposition 3.1.2 we know the form of the $s \times s$ block matrices $K = (K_{ij})$ that centralize A . Let's

count the number of independent choices we have for the entries of K regarded as an unblocked $n \times n$ matrix. That will be the dimension of the centralizer. (We could write down a basis but we will avoid formalities.) The entries of any one of the $m_i \times m_j$ matrices K_{ij} can be chosen independently of those for another. Also the number of choices for the entries of K_{ij} is the smaller of m_i, m_j , that is, the “ m ” corresponding to the bigger index of i or j . For a given $1 \leq k \leq s$ there are $2k - 1$ pairs of indices (i, j) for which $\max\{i, j\} = k$. For each such pair, the corresponding K_{ij} can have m_k independent choices of NW-SE diagonals. Thus, the total number of independent choices for entries in the K_{ij} with $\max\{i, j\} = k$ is $(2k - 1)m_k$. Hence, the number of free choices for the entries of K is

$$\dim \mathcal{C}(A) = m_1 + 3m_2 + \cdots + (2k - 1)m_k + \cdots + (2s - 1)m_s. \quad \square$$

3.2 THE CENTRALIZER OF A WEYR MATRIX

In Chapter 2, Proposition 2.3.3, we described the matrices that centralize a given nilpotent matrix in Weyr form. For the reader’s convenience, we restate that description here.

Proposition 3.2.1

Let W be an $n \times n$ nilpotent Weyr matrix with Weyr structure (n_1, \dots, n_r) , where $r \geq 2$. Let K be an $n \times n$ matrix, blocked according to $n_i \times n_i$ diagonal blocks, and let K_{ij} denote its (i, j) block for $i, j = 1, \dots, r$. Then W and K commute if and only if K is a block upper triangular matrix for which

$$K_{ij} = \begin{bmatrix} K_{i+1,j+1} & * \\ 0 & * \end{bmatrix} \text{ for } 1 \leq i \leq j \leq r - 1,$$

*where the column of asterisks disappears if $n_j = n_{j+1}$ and the $[0 \ *]$ row disappears if $n_i = n_{i+1}$.*

We now calculate the dimension of $\mathcal{C}(A)$ for a nilpotent matrix A in terms of its Weyr structure. The resulting formula makes an interesting contrast to the Frobenius formula, which we derived in the previous section. The new formula does not appear to be widely known.²

2. Nor does there appear to be a name attached to this formula. The authors do not know who first derived the formula by a direct matrix argument, as compared with deducing it from the Frobenius formula using duality and the combinatorial result in our later Proposition 3.3.1.

Proposition 3.2.2

Let A be a nilpotent $n \times n$ matrix over a field F , and let (n_1, n_2, \dots, n_r) be the Weyr structure of A . Then

$$\dim \mathcal{C}(A) = n_1^2 + n_2^2 + \dots + n_r^2.$$

Proof

We can assume A is already in Weyr form because a similarity change won't alter the dimension of the centralizer. By Proposition 3.2.1, the matrices K that centralize A are exactly the $r \times r$ block upper triangular matrices (with the same block structure as A)

$$K = \begin{bmatrix} K_{11} & K_{12} & K_{13} & \cdots & K_{1r} \\ 0 & K_{22} & K_{23} & \cdots & K_{2r} \\ 0 & 0 & K_{33} & \cdots & K_{3r} \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & 0 & K_{rr} \end{bmatrix}$$

for which

$$K_{ij} = \begin{bmatrix} K_{i+1,j+1} & * \\ 0 & * \end{bmatrix} \text{ for } 1 \leq i \leq j \leq r-1.$$

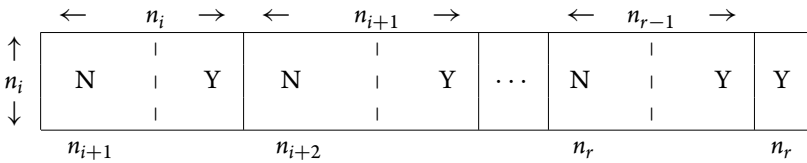
Let's count the number of independent choices we have for the entries of such a matrix as an $n \times n$ matrix, starting with the bottom row of blocks.

Claim: As we progress up the rows of blocks, each i th row gives us exactly an additional free choice of n_i^2 entries.

We argue recursively. Clearly, we have a free choice of n_r^2 for the entries of the $n_r \times n_r$ matrix K_{rr} , and so for the last row of blocks. Now fix $1 \leq i < r$ and suppose we have chosen the entries for the last $r-i$ rows of blocks. By Proposition 3.2.1, our choices for the entries in the i th row of blocks are then constrained only by the relationship above between K_{ij} and $K_{i+1,j+1}$ for $i \leq j \leq r-1$. That is, we can only freely choose the asterisk entries of

$$K_{ij} = \begin{bmatrix} K_{i+1,j+1} & * \\ 0 & * \end{bmatrix}.$$

In terms of the picture (of the nonzero blocks in row i),



we can freely choose the Y (yes) parts but have no choice in the N (no) parts. This provides us an additional free choice of

$$n_i [(n_i - n_{i+1}) + (n_{i+1} - n_{i+2}) + \cdots + (n_{r-1} - n_r) + n_r] = n_i^2$$

entries in the i th row of blocks. Thus, the induction works and the claim is verified. Therefore, we have a total of $n_r^2 + n_{r-1}^2 + \cdots + n_2^2 + n_1^2$ independent choices for the entries of the matrix K , giving the dimension of $\mathcal{C}(A)$ as claimed. (We could formalize this argument by explicitly exhibiting a basis, but this seems unnecessary.) □

The following example illustrates our two contrasting descriptions of the centralizer of a nilpotent matrix and the centralizer dimension.

Example 3.2.3

Let A be a nilpotent 7×7 matrix with Jordan structure $(3, 2, 2)$ and therefore dual Weyr structure $(3, 3, 1)$ by Theorem 2.4.1. If we put A in Jordan form, then by Proposition 3.1.2 a typical matrix K that centralizes A looks like

$$K = \left[\begin{array}{ccc|ccc} a & b & c & d & e & f & g \\ 0 & a & b & 0 & d & 0 & f \\ 0 & 0 & a & 0 & 0 & 0 & 0 \\ \hline 0 & h & i & j & k & l & m \\ 0 & 0 & h & 0 & j & 0 & l \\ \hline 0 & n & p & q & r & s & t \\ 0 & 0 & n & 0 & q & 0 & s \end{array} \right].$$

Computing $\dim \mathcal{C}(A)$ according to the Frobenius formula in Proposition 3.1.3, we have

$$\dim \mathcal{C}(A) = 1 \times 3 + 3 \times 2 + 5 \times 2 = 19.$$

That is consistent with our use of 19 letters to label the nonzero entries in a typical centralizing matrix K .

Now suppose we put A in Weyr form. Its Weyr structure is $(3, 3, 1)$ and, by Proposition 3.2.1, a typical centralizing matrix³ K looks like

$$K = \left[\begin{array}{ccc|ccc|c} a & b & c & h & k & l & r \\ 0 & d & e & i & m & n & s \\ 0 & f & g & j & p & q & t \\ \hline & & & a & b & c & h \\ & & & 0 & d & e & i \\ & & & 0 & f & g & j \\ \hline & & & & & & a \end{array} \right].$$

By our Weyr form formula in Proposition 3.2.2, we get

$$\dim \mathcal{C}(A) = 3^2 + 3^2 + 1^2 = 19.$$

Again the answer matches the number of letters used to label the nonzero entries of a typical centralizing matrix using the Weyr form description.⁴ \square

We return to some unfinished business in Chapter 1, namely the proof of Proposition 1.1.2, which gave equivalent conditions for a matrix to be nonderogatory. Recall that, over an algebraically closed field F , a matrix $A \in M_n(F)$ is called **nonderogatory** if all of its eigenspaces are 1-dimensional. The proof we now give uses the Jordan form and the Frobenius formula (but we could equally as well have used the Weyr formula instead). It is yet another instance of the convenience of being able to reduce a matrix to a canonical form. A purist might argue that it is not necessary to go all the way to a canonical form for the proof (see Remark 3.2.5), but we already have this machinery assembled and so may as well use it to give a shorter proof than the usual one. We also augment Proposition 1.1.2 with another three useful equivalent conditions.

Proposition 3.2.4

The following are equivalent for an $n \times n$ matrix A over an algebraically closed field F :

- (1) A is nonderogatory.
- (2) $\dim F[A] = n$.
- (3) $\dim \mathcal{C}(A) = n$.
- (4) $\mathcal{C}(A) = F[A]$. In other words, the only matrices that commute with A are polynomials in A .

3. The Jordan and Weyr forms correspond under a permutation similarity transformation (Theorem 2.4.1), and this induces an isomorphism of the centralizers. But this “second K ” is not the image of the “first K ” under this isomorphism. Subtle point. What is the correct image?

4. It is also reassuring to get the same dimension in both the Jordan and Weyr calculations!

- (5) In the Jordan form of A , there is only one basic Jordan block for each eigenvalue of A .
- (6) In the Weyr form of A , the Weyr structures associated with the eigenvalues of A are $(1, 1, \dots, 1)$.

Proof

Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be the distinct eigenvalues of A . By the Corollary 1.5.4 to the generalized eigenspace decomposition, A is similar to a matrix $B = \text{diag}(B_1, B_2, \dots, B_k)$, where B_i has λ_i as its sole eigenvalue for $i = 1, 2, \dots, k$. Similar matrices have the same Jordan form and the same Weyr form. Since similar matrices also have isomorphic centralizers, generate isomorphic subalgebras, and share or fail the nonderogatory property together, clearly it suffices to establish the proposition for B . Now B is nonderogatory precisely when each B_i is nonderogatory. Note that $\dim F[B] = \sum_{i=1}^k \dim F[B_i]$ because the minimal polynomial of B is the product of the minimal polynomials of the B_i (these being relatively prime) and for any matrix M , we know from Proposition 1.4.2 that $\dim F[M]$ agrees with the degree of the minimal polynomial of M . (Alternatively, we can deduce the dimension equation from the fact, to be established in Proposition 5.1.1, that $F[B] \cong \prod_{i=1}^k F[B_i]$.) Also $\mathcal{C}(B) \cong \prod_{i=1}^k \mathcal{C}(B_i)$ by Proposition 3.1.1, whence $\dim \mathcal{C}(B) = \sum_{i=1}^k \dim \mathcal{C}(B_i)$. Consequently, it is enough to get the proposition for each B_i . The upshot of our reductions is that, without loss of generality, we can assume that A has a single eigenvalue λ and is a Jordan matrix, say with Jordan structure (m_1, m_2, \dots, m_s) .

Noting that $\mathcal{C}(\lambda I + N) = \mathcal{C}(N)$ for any $N \in M_n(F)$, by the Frobenius formula 3.1.3 applied to the nilpotent $N = A - \lambda I$ we have

$$(i) \quad \dim \mathcal{C}(A) = m_1 + 3m_2 + 5m_3 + \dots + (2s-1)m_s.$$

Also, since $\dim F[A]$ agrees with the degree of the minimal polynomial of A , and the minimal polynomial of our Jordan matrix A is $(x - \lambda)^{m_1}$, we know that

$$(ii) \quad \dim F[A] = m_1.$$

Inasmuch as $F[A] \subseteq \mathcal{C}(A)$, we have $F[A] = \mathcal{C}(A)$ if and only if $\dim F[A] = \dim \mathcal{C}(A)$. Moreover, since $n = m_1 + m_2 + \dots + m_s$, clearly we have $\dim \mathcal{C}(A) \geq n$ by (i), and $\dim F[A] \leq n$ by (ii). It follows that (4) is equivalent to both (2) and (3) holding at once. Also (5) and (6) are equivalent because the Weyr structure associated with an eigenvalue is the dual of the Jordan structure by Theorem 2.4.1. So to complete the proof, we need only show that (1), (2), and (3) are each equivalent to (5), the latter saying $s = 1$ because of our simplified assumptions.

The equivalence of (1) and (5) is clear. (More generally, the dimension of the eigenspace $E(\lambda_i)$ is the number of basic Jordan blocks corresponding to λ_i .) Also by (ii), and the fact that $n = m_1 + m_2 + \dots + m_s$, we see that $\dim F[A] = n$

if and only if $s = 1$. Thus, (2) and (5) are equivalent. Finally, by (i), we see that (3) holds if and only if $m_1 + 3m_2 + 5m_3 + \cdots + (2s - 1)m_s = m_1 + m_2 + m_3 + \cdots + m_s$, again equivalent to $s = 1$. So (3) and (5) are equivalent also. \square

Remarks 3.2.5

- (1) With our definition of nonderogatory, Proposition 3.2.4 fails over a field that is not algebraically closed. For this reason, some authors prefer to define a nonderogatory matrix $A \in M_n(F)$ as one that possesses a **cyclic vector**, that is, a vector $v \in F^n$ for which $\{v, Av, A^2v, \dots, A^{n-1}v\}$ is a basis for F^n . This is equivalent to A being similar to a companion matrix (see Example 1.4.3). With the new definition, the equivalence of (1) to (4) holds over any field F . However, one really has to call upon the (nontrivial) cyclic decomposition of F^n into a direct sum of A -cyclic subspaces to establish this. (See Chapter 7 of the Hoffman and Kunze text *Linear Algebra*, and also our Section 4.6 in Chapter 4.) On the other hand, some of the implications are easy (and neat). For instance, to show that (1) \Rightarrow (4) we argue that if v is a cyclic vector for A , and B commutes with A , then we can write

$$Bv = a_0v + a_1Av + a_2A^2v + \cdots + a_{n-1}A^{n-1}v$$

for some scalars a_i , so letting

$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} \in F[x]$, we have

$$BA^i v = A^i Bv = A^i p(A)v = p(A)A^i v \quad \text{for } i = 0, 1, \dots, n-1.$$

Therefore, $B = p(A)$ because $\{v, Av, A^2v, \dots, A^{n-1}v\}$ is a basis. However, our interest in nonderogatory matrices, more generally k -regular matrices, lies in the geometric multiplicity of the eigenvalues, not cyclic vectors.

- (2) Other authors define a matrix $A \in M_n(F)$ to be nonderogatory if its minimal polynomial $m(x)$ is equal to its characteristic polynomial $p(x)$. Since this is equivalent to saying that $\deg m(x) = n$, it follows quickly from the equivalence of (1) and (2) of Proposition 3.2.4 that the two definitions agree. \square

3.3 A MATRIX STRUCTURE INSIGHT INTO A NUMBER-THEORETIC IDENTITY

This short section is for light relief, not to be taken too seriously, particularly the “philosophizing.” We have, after all, been climbing steadily for some time now and deserve a break.

As we all know, the sum of the first s odd numbers is a square:

$$1 + 3 + 5 + \cdots + (2s - 1) = s^2.$$

Much less well known is the following generalization.

Proposition 3.3.1

Let (m_1, m_2, \dots, m_s) be a partition of a positive integer n . Let (n_1, n_2, \dots, n_r) be the dual (or conjugate) partition of (m_1, m_2, \dots, m_s) . Then

$$m_1 + 3m_2 + 5m_3 + \cdots + (2s - 1)m_s = n_1^2 + n_2^2 + \cdots + n_r^2.$$

So, for example, $(3, 3, 2, 1, 1)$ and $(5, 3, 2)$ are dual partitions of 10, whence

$$3 + 3 \times 3 + 5 \times 2 + 7 \times 1 + 9 \times 1 = 5^2 + 3^2 + 2^2.$$

The classical special case (above) comes from taking the partition $(1, 1, \dots, 1)$ of $n = s$ (so all $m_i = 1$). The dual partition is (n) (so $r = 1$ and $n_1 = s$). The general identity is not deep—it basically follows from repeated applications of the classical one. Gerstenhaber did include a short induction proof of the identity of Proposition 3.3.1 (but without using matrix connections⁵) in his 1961 paper. That he chose to do so, in no less a journal than the *Annals of Mathematics*, suggests the identity is not entirely trivial!

Proof of 3.3.1. The identity follows easily (and naturally) from our two formulae for calculating the dimension of the centralizer of a nilpotent matrix. For let A be an $n \times n$ nilpotent matrix with Jordan structure (m_1, m_2, \dots, m_s) . By Theorem 2.4.1, A has for its Weyr structure the dual partition (n_1, n_2, \dots, n_r) . Therefore, by Propositions 3.1.3 and 3.2.2, we have

$$m_1 + 3m_2 + 5m_3 + \cdots + (2s - 1)m_s = \dim \mathcal{C}(A) = n_1^2 + n_2^2 + \cdots + n_r^2. \quad \square$$

Of course, if one had suspected the identity in 3.3.1, a neat way of confirming it would be to use one of the oldest tricks in the combinatorial book—equating row sums of some array to column sums. Specifically, we fill in the entries of the

5. Gerstenhaber's interest in the identity was, however, in connection with the Frobenius Formula 3.1.3 for the dimension of the centralizer of a nilpotent Jordan matrix. One assumes he was unaware of the corresponding formula 3.2.2 in terms of Weyr structure, or possibly, for that matter, of the Weyr form itself.

Young tableau for the partition (m_1, m_2, \dots, m_s) with 1's in the first row, 3's in the second row, 5's in the third row, and so on:

1	1	1
3	3	3
5	5	
7		
9		

(in the case of the partition $(3, 3, 2, 1, 1)$ above). The sum of the entries over all the rows is now the left side of the identity. Using the classical identity (and the very nature of a dual partition), we see that the column sums are

$$\begin{aligned}
 1 + 3 + 5 + \cdots + (2n_1 - 1) &= n_1^2 \\
 1 + 3 + 5 + \cdots + (2n_2 - 1) &= n_2^2 \\
 &\vdots \\
 1 + 3 + 5 + \cdots + (2n_r - 1) &= n_r^2.
 \end{aligned}$$

Hence, the total column sums give the right side of the identity.

By a similar argument, one could produce all sorts of other identities by starting with a known identity and filling in the entries of the Young tableau accordingly. However, such identities would not be as pretty as 3.3.1 (and would look contrived).

Some would say that the above combinatorial proof of 3.3.1 has a pedagogical weakness, namely, it requires a “trick” (and maybe even to know the answer in advance). The matrix centralizer approach, on the other hand, leads one to the answer and provides a meaningful interpretation of the common number in the two sides of the identity.⁶ Curiously, the matrix argument does not (overtly) make use of the classical identity.⁷

Before resuming our real work, this seems a good opportunity to impart a little mathematical philosophy. There is a tendency in some quarters to dismiss a result in mathematics (such as 3.3.1) as not being that important if the proof is very simple. That philosophy can easily be rebutted by two of the most useful and beautiful results in finite group theory: Lagrange’s theorem (the order

6. The authors were unaware of the identity prior to making the matrix connections.

7. Tom Roby of the University of Connecticut has asked (privately) whether there is also a group-theoretic insight into the identity 3.3.1. The right-hand side certainly reminds one of “the sum of the squares of the degrees of the irreducible representations (say over \mathbb{C}) of a finite group equals the order of the group.”

of a subgroup divides the order of the group) and Burnside's orbit-counting theorem. The latter theorem (also proved by equating certain row sums to column sums) enables even those with meager geometric visual skills to count, for example, the number of "different" ways (allowing for rotations) of coloring the sides of a regular solid, given a fixed number of colors. Or to count the number of different chemical compounds that can be obtained by attaching given radicals to a given molecule at specified atoms. The reader will notice a number of results on the Weyr form, throughout our book, that are also pretty easy to prove, such as Theorem 2.4.1, Proposition 3.2.2, and Theorem 3.4.3 of the next section. Don't underrate their importance!

3.4 LEADING EDGE SUBSPACES OF A SUBALGEBRA

And now for some leading edge technology.

Suppose W is a fixed nilpotent $n \times n$ matrix in Weyr form and with Weyr structure (n_1, n_2, \dots, n_r) . By Proposition 3.2.1 we know, as blocked matrices of the same block structure as W , the precise form of $r \times r$ matrices $K = (K_{ij})$ that centralize W . **Such a matrix is completely determined by its top row of blocks** $[K_{11}, K_{12}, \dots, K_{1r}]$. In the interest of notational simplicity, we often represent K by its first row of blocks, providing the context and underlying Weyr structure are clear. (Normally, we will warn the reader when we are using this **top row notation**.) Thus, in top row notation $[X_1, X_2, \dots, X_r]$ is the $r \times r$ block matrix in $\mathcal{C}(W)$ whose first row has the blocks X_1, X_2, \dots, X_r . Note that X_i is an $n_1 \times n_i$ matrix. For instance, if the Weyr structure of W is $(3, 2, 2)$, then in top row notation,

$$K = \begin{array}{|ccc|ccc|cc|} \hline 1 & 1 & 4 & 4 & 6 & 2 & 1 \\ 2 & 3 & 6 & 7 & 5 & 8 & 6 \\ 0 & 0 & 9 & 0 & 0 & 3 & 9 \\ \hline \end{array}$$

could only be (by Proposition 3.2.1) the 7×7 centralizing matrix

$$K = \left[\begin{array}{ccc|cc|cc} 1 & 1 & 4 & 4 & 6 & 2 & 1 \\ 2 & 3 & 6 & 7 & 5 & 8 & 6 \\ 0 & 0 & 9 & 0 & 0 & 3 & 9 \\ \hline & & & 1 & 1 & 4 & 6 \\ & & & 2 & 3 & 7 & 5 \\ \hline & & & & & 1 & 1 \\ & & & & & 2 & 3 \end{array} \right].$$

Warning. We must take care to choose only top rows that do actually correspond to the top row of some matrix that centralizes W . For instance, in the above example, we couldn't start with a top row having a nonzero $(3, 1)$ entry in the first block of K .

A corresponding shorthand for matrices that centralize a nilpotent Jordan matrix would be very awkward, suggesting yet again that matrices in Weyr form are often easier to deal with in commuting problems.

Staying with our fixed matrix W , suppose \mathcal{A} is a subalgebra of $\mathcal{C}(W)$. The case of principal interest later is when \mathcal{A} is a commutative subalgebra of $M_n(F)$ containing W . We can associate with \mathcal{A} the following “leading edge subspaces” of $n_1 \times n_{i+1}$ matrices.

Definition 3.4.1: Let W be a nilpotent Weyr matrix with Weyr structure (n_1, n_2, \dots, n_r) , and let \mathcal{A} be a subalgebra of $\mathcal{C}(W)$. For $i = 0, 1, \dots, r - 1$, let

$$U_i(\mathcal{A}) = \{X \in M_{n_1 \times n_{i+1}}(F) : [0, 0, \dots, 0, X, *, *, \dots, *] \in \mathcal{A}\}$$

where in the top row notation we have used, the string of zero blocks is of length i , and the stars $*$ represent unspecified entries (i.e., for some choice of $*$, not all). We call $U_0(\mathcal{A}), U_1(\mathcal{A}), \dots, U_{r-1}(\mathcal{A})$ the **leading edge subspaces associated with \mathcal{A}** (relative to W). If \mathcal{A} is understood, we write these subspaces more simply as U_0, U_1, \dots, U_{r-1} .

The reader could very sensibly ask why we didn't index the leading edge subspaces by U_1, U_2, \dots, U_r so that the index i matches that in the number n_i of columns of matrices in U_i ? The rationale for going with the above indexing is twofold: (1) to correspond with the nonzero powers $W^0 (= I), W^1, \dots, W^{r-1}$ of the Weyr matrix W (W has nilpotency index r , and if $W \in \mathcal{A}$ then the power W^i contributes to our U_i for the range $i = 0, 1, \dots, r - 1$), and (2) to get nice expressions of properties such as $U_i U_j \subseteq U_{i+j}$ (see Proposition 3.4.4 below). The latter property in the alternative notation would be the unfriendly $U_i U_j \subseteq U_{i+j-1}$. Our adopted notation does have drawbacks at times, but we feel the benefits outweigh the disadvantages. To help the reader digest these ideas and notation, we now serve up a little example.

Example 3.4.2

Let \mathcal{A} be the subalgebra of $M_4(F)$ generated by the pair of matrices

$$W = \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & 1 \\ & & & 0 \end{array} \right], \quad K = \left[\begin{array}{cc|cc} 2 & 1 & 1 & 1 \\ 0 & 1 & 0 & 3 \\ \hline & & 2 & 1 \\ & & & 2 \end{array} \right].$$

Thus, \mathcal{A} consists of all linear combinations of products involving I , W , and K (such as $5I + 3K^2W - 2KW^2$). Note that W is a nilpotent Weyr matrix of structure $(2, 1, 1)$. Just by observing the form of K , we see that W and K commute because of Proposition 3.2.1, so \mathcal{A} is a commutative subalgebra containing W . There are three leading edge subspaces of \mathcal{A} relative to W , namely U_0 , U_1 , U_2 . Certainly

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \in U_0$$

because these matrices are the leading $(1, 1)$ blocks of subalgebra members I_4 and K , respectively. Similarly, because

$$K^2 = \left[\begin{array}{cc|cc} 4 & 3 & 4 & 8 \\ 0 & 1 & 0 & 9 \\ \hline & & 4 & 4 \\ & & \hline & & & 4 \end{array} \right], K^2W = \left[\begin{array}{cc|cc} 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & 4 \\ & & \hline & & & 0 \end{array} \right], W^2 = \left[\begin{array}{cc|cc} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & 0 \\ & & \hline & & & 0 \end{array} \right]$$

are all members of \mathcal{A} , we see from their leading first row blocks that

$$\begin{bmatrix} 4 & 3 \\ 0 & 1 \end{bmatrix} \in U_0, \begin{bmatrix} 4 \\ 0 \end{bmatrix} \in U_1, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \in U_2.$$

On the other hand, it would be a mistake to look at the $(1, 3)$ block of K^2 and conclude that

$$\begin{bmatrix} 8 \\ 9 \end{bmatrix} \in U_2.$$

It will follow easily from results to come that, with this subalgebra \mathcal{A} , the dimensions of the three leading edge subspaces are respectively 2, 1, and 1, and $\dim \mathcal{A} = \dim U_0 + \dim U_1 + \dim U_2 = 4$. In fact, U_0 is spanned by the first displayed pair, while U_1 and U_2 are spanned by their respectively singled-out member. \square

It should be clear to the reader why we have used the term “leading edge subspace”—we have picked out the leading blocks X , at a given distance in, from the first rows of contributing matrices in our algebra \mathcal{A} . In the full matrix picture, the corresponding superdiagonal of such a contributing matrix is a foremost wing edge, with essentially repeated blocks of X . (Think of an aircraft wing.) See Figure 3.1, which illustrates the homogeneous case.

But why bother with these spaces? Part of the answer lies in the following very useful formula.

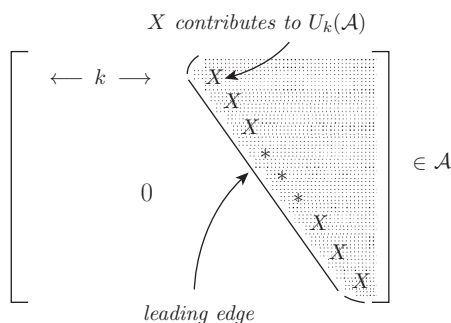


Figure 3.1 Leading edge subspace $U_k(\mathcal{A})$

Theorem 3.4.3 (The Leading Edge Dimension Formula)

In the notation of Definition 3.4.1,

$$\dim \mathcal{A} = \dim U_0 + \dim U_1 + \cdots + \dim U_{r-1}.$$

Proof

Throughout the proof, we will treat our matrices as blocked matrices having the same block structure as W . It is convenient to work inside the algebra \mathcal{T} of all $r \times r$ block upper triangular matrices. Note that $\mathcal{A} \subseteq \mathcal{C}(W) \subseteq \mathcal{T}$, where the first inclusion is by assumption and the second by Proposition 3.2.1. For $i = 1, 2, \dots, r-1$, let

$$\pi_i : \mathcal{T} \rightarrow \mathcal{T}$$

be the projection of \mathcal{T} onto its top $i \times i$ left-hand corner of blocks:

$$\begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1r} \\ 0 & X_{22} & X_{23} & \cdots & X_{2r} \\ 0 & 0 & X_{33} & \cdots & X_{3r} \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & 0 & X_{rr} \end{bmatrix} \mapsto \begin{bmatrix} X_{11} & \cdots & X_{1i} & 0 & \cdots \\ \vdots & & \vdots & \vdots & \\ 0 & \cdots & X_{ii} & 0 & \cdots \\ 0 & \cdots & 0 & 0 & \cdots \\ \vdots & & & & \end{bmatrix}$$

This is an algebra homomorphism,⁸ although for the present proof we only need the fact that it is a linear transformation. (In later chapters, we will use the full algebra homomorphism properties of these lovely maps when working with the Weyr form, the Jordan analogues of which are nowhere nearly as suggestive.) Let $\mathcal{A}_i = \pi_i(\mathcal{A})$.

8. Algebra homomorphisms are required to preserve sums, products, and scalar multiplication. The authors would also (ideally) like the homomorphism to map the identity to the identity, although that doesn't matter here.

Let us first consider π_{r-1} . When restricted to \mathcal{A} , its kernel is naturally isomorphic (as a vector space) to the leading edge subspace U_{r-1} . (Remember that matrices centralizing W are completely determined by their top row of blocks.) Thus,

$$\dim \mathcal{A} = \dim \mathcal{A}_{r-1} + \dim U_{r-1}.$$

Next consider π_{r-2} . When restricted to \mathcal{A}_{r-1} , its kernel is isomorphic to U_{r-2} and so $\dim \mathcal{A}_{r-1} = \dim \mathcal{A}_{r-2} + \dim U_{r-2}$. Therefore,

$$\dim \mathcal{A} = \dim \mathcal{A}_{r-2} + \dim U_{r-2} + \dim U_{r-1}.$$

Continuing down this path by successively applying the projections $\pi_{r-1}, \pi_{r-2}, \dots, \pi_1$ leads to

$$\dim \mathcal{A} = \dim \mathcal{A}_1 + \dim U_1 + \dim U_2 + \dots + \dim U_{r-1}.$$

But \mathcal{A}_1 is isomorphic to U_0 , so $\dim \mathcal{A}_1 = \dim U_0$ and our proof is complete.

We remark in passing that, since the algebra \mathcal{A} is vector space isomorphic to its space of top rows of blocks, we could have done the projection arguments on just the latter vector space. We feel, however, that this is a good place to introduce the corner projection arguments. \square

In general, although there are often ways of bounding the dimension of an individual leading edge subspace (some are mentioned in Chapter 5), the dimensions of the leading edge subspaces don't bear much relationship to each other except when W belongs to \mathcal{A} and some $n_i = n_{i+1}$. (See Example 3.5.2.) We record the following leading edge properties for future use. Two of the statements use “centralize” applied not to individual matrices but to sets of matrices. We say that a set $V \subseteq M_n(F)$ **centralizes** the set $U \subseteq M_n(F)$ if all matrices in V commute with everything in U . And a subalgebra U of $M_n(F)$ is said to be **self-centralizing**⁹ if U is commutative but no set $V \subseteq M_n(F)$ that properly contains U centralizes U . Self-centralizing subalgebras will be studied in more detail in Chapter 5.

Proposition 3.4.4

Let \mathcal{A} be a commutative subalgebra of $M_n(F)$ containing a nilpotent Weyr matrix W of Weyr structure (n_1, n_2, \dots, n_r) . Let U_0, U_1, \dots, U_{r-1} be the leading edge subspaces of \mathcal{A} relative to W . Then:

- (1) U_0 is a commutative subalgebra of $M_{n_1}(F)$.
- (2) $\dim U_i \geq 1$ for all i .

9. Being a self-centralizing subalgebra is equivalent to being a maximal commutative subalgebra.

- (3) If $n_i = n_{i+1}$ for some i , then $U_{i-1} \subseteq U_i$. In particular, in this case, $\dim U_{i-1} \leq \dim U_i$.
- (4) If the Weyr structure is homogeneous ($n_1 = n_2 = \dots = n_r$), then U_i centralizes U_j whenever $i + j < r$.
- (5) In the homogeneous case, $U_i U_j \subseteq U_{i+j}$ whenever $i + j < r$.
- (6) In the homogeneous case, if U_0 is a self-centralizing subalgebra of $M_{n_1}(F)$ of dimension d , then $\dim \mathcal{A} = dr$.

Proof

These properties are easy to establish. Let $X \in U_i$ and $Y \in U_j$ where $i + j < r$. In top row notation, suppose $K = [0, 0, \dots, 0, X, *, *, \dots, *] \in \mathcal{A}$ (string of i zeros) and $L = [0, 0, \dots, 0, Y, *, *, \dots, *] \in \mathcal{A}$ (string of j zeros). Then in the homogeneous case,

$$KL = [0, 0, \dots, 0, XY, *, *, \dots, *]$$

with a string of $i + j$ zeros. Similarly in top row notation, we have $LK = [0, 0, \dots, 0, YX, *, *, \dots, *]$ (same string of zeros), whence $XY = YX$ because \mathcal{A} is commutative. This gives (4) and (5). A slight modification of the argument gives (1). (In fact, (4) and (5) hold more generally in the nonhomogeneous case provided also $n_1 = n_{i+j+1}$.) For (2), observe that for $i = 0, 1, \dots, r - 1$, the power $W^i \in \mathcal{A}$ contributes to U_i the $n_1 \times n_{i+1}$ matrix

$$\begin{bmatrix} I \\ 0 \end{bmatrix}.$$

Now we establish (3). Because of the shifting effect on columns of blocks under right multiplication by W (see Remark 2.3.1), in a product KW with $K \in \mathcal{A}$, the i th block in the first row of K is faithfully shifted one place to its right because $n_{i+1} = n_i$. Therefore, if $X \in U_{i-1}$ comes from the matrix $K \in \mathcal{A}$, then X can also be viewed as the leading block $i + 1$ steps inside, along the first row of blocks of the matrix $KW \in \mathcal{A}$, placing $X \in U_i$. Thus, $U_{i-1} \subseteq U_i$.

Finally, let's look at (6), which is the most interesting of the properties. Assume U_0 is self-centralizing, which implies that an $n_1 \times n_1$ matrix that commutes with all matrices in U_0 must already be in U_0 . Then by (3) and (4), we must have $U_0 = U_1 = \dots = U_{r-1}$. By Theorem 3.4.3,

$$\begin{aligned} \dim \mathcal{A} &= \dim U_0 + \dim U_1 + \dots + \dim U_{r-1} \\ &= d + d + \dots + d \\ &= dr. \end{aligned}$$

What could be simpler than that! The Weyr form often suggests neat arguments. □

We will illustrate the utility of some of the above properties in the next section.

3.5 COMPUTING THE DIMENSION OF A COMMUTATIVE SUBALGEBRA

The subject matter here will be covered in much more detail in Chapter 5, and without loss of continuity, the reader can choose to skip this present section. However, we feel it is to the reader's benefit to work through some numerical examples at this stage, involving the leading edge subspaces and the Leading Edge Dimension Formula 3.4.3.

Suppose \mathcal{A} is a commutative subalgebra of $M_n(F)$. What are the possibilities for the (vector space) dimension of \mathcal{A} ? At the outset, we must say that in general this is a very difficult problem. Schur in 1905 provided the sharp upper bound of $\lfloor n^2/4 \rfloor + 1$ for the dimension,¹⁰ where $\lfloor y \rfloor$ indicates the largest integer less than or equal to a real number y . (The bound can be realized when n is even by the commutative subalgebra of all $n \times n$ matrices having a scalar diagonal, arbitrary entries in the top right $n/2 \times n/2$ corner, and zeros elsewhere.) But even if \mathcal{A} is generated as a subalgebra by three commuting $n \times n$ matrices A, B, C , it is not known for instance whether the dimension of \mathcal{A} can exceed n . In this section, we look at three numerical examples to illustrate how information on the leading edge subspaces of \mathcal{A} can aid the calculation of $\dim \mathcal{A}$. The calculations can be done by hand, but Matlab or Maple can also be put on standby for the numerically challenged.¹¹

Example 3.5.1

The following are three 9×9 commuting matrices, the first of which is nilpotent of nilpotency index 3:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & -3 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix},$$

10. For a very short ring-theoretic proof of this, see Cowsik's 1993 paper.

11. It is said that one doesn't really fully understand an algorithm unless one can do the calculations by hand in a moderately sized example.

$$B = \begin{bmatrix} 1 & 6 & 0 & 0 & 0 & 3 & -9 & 3 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 6 & 2 & 0 \\ 0 & 0 & -1 & 6 & 6 & 0 & 0 & 0 & 2 \\ 0 & 0 & -1 & 4 & 3 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 4 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -3 & 0 & 0 \\ 0 & 0 & 1 & -3 & -3 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$C = \begin{bmatrix} 4 & -3 & 0 & 0 & 0 & 0 & 9 & 3 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 3 & -1 & 0 \\ 0 & -9 & 2 & 3 & 3 & 9 & 27 & -9 & -1 \\ 0 & -3 & 0 & 4 & 3 & 3 & 9 & -6 & 0 \\ 0 & -3 & 0 & 0 & 1 & 0 & 9 & 3 & 0 \\ 0 & 6 & 0 & 0 & 0 & 4 & -9 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 & 3 & 3 & 0 \\ 0 & 0 & -2 & 3 & 3 & 0 & 0 & 9 & 3 \end{bmatrix}.$$

What is the dimension of the subalgebra $\mathcal{A} = F[A, B, C]$ of $M_9(F)$ generated by A , B , and C ? A brute force method might proceed as follows. By the Cayley–Hamilton theorem and the fact that A, B, C commute, \mathcal{A} is spanned as a vector space by the matrices $A^i B^j C^k$ where $0 \leq i, j, k \leq 8$. (We could somewhat restrict this range of indices if we had more information on the degrees of the minimal polynomials of A, B, C .) Write each of these 9^3 matrices as a $9^2 \times 1$ column vector (by running through the matrix entries in some fixed order), and then form the $9^2 \times 9^3$ matrix M whose columns are the said column vectors. Then $\dim \mathcal{A}$ is the rank of this 81×729 matrix M ! The rank calculation of M (say by row operations) would really test one's hand calculations, so let's get smarter.

Firstly we put A in Weyr form, following the algorithm in Section 2.5 of Chapter 2. The reader who does these calculations should finish up with the same Weyr matrix (or else one of us is wrong) but he or she may have used a different similarity transformation to the one the authors arrived at. (One has choices in the way one row reduces a matrix, establishes a basis of a null space, and extends a basis.) Let $A_1 = A$. By elementary row operations we can determine a basis for the null space of A_1 , and extend this to a basis for F^9 . (A few row operations on A_1 reveal its rank is 6 and so its nullity is 3. One can pick out a simple basis for the null space by just eye-balling A_1 .) Under the resulting change of basis, A_1 is transformed to a matrix of

the form

$$P_1^{-1}A_1P_1 = \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \\ \hline 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \end{array} \right],$$

where B_2 is 3×6 and A_2 is 6×6 . Next we perform the same procedure on the bottom corner matrix A_2 by finding a basis for its null space and extending it to a basis for F^6 . (Again the nullity is 3 and a natural choice presents itself here for a basis and an extension.) Suppose transforming A_2 under the resulting change of basis is achieved via conjugation by the 6×6 invertible matrix P_2 . Then we conjugate $P_1^{-1}A_1P_1$ by $\text{diag}(I_3, P_2)$, which turns out to yield a strictly block upper triangular matrix X with 3×3 blocks

$$X = \begin{bmatrix} 0 & X_{12} & X_{13} \\ 0 & 0 & X_{23} \\ 0 & 0 & 0 \end{bmatrix}.$$

The algorithm now tells us that A has a homogeneous Weyr structure $(3, 3, 3)$. We need to convert X_{12} and X_{23} to identity blocks, and X_{13} to a zero block. Necessarily, X_{12} and X_{23} are rank 3 matrices, hence invertible 3×3 matrices. Conjugating X by $\text{diag}(X_{12}, I_3, X_{23}^{-1})$ converts the $(1, 2)$ and $(2, 3)$ blocks of X to identity matrices, and a further conjugation in the form of elementary row operations clears out the $(1, 3)$ block to yield a matrix in Weyr form. The net result in the authors' particular calculations was that conjugation by the matrix

$$D = \frac{1}{3} \begin{bmatrix} 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 1 \\ 6 & 3 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 1 \\ -3 & 3 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

puts A in the Weyr form

$$W = D^{-1}AD = \left[\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 1 & 0 & 0 & & & \\ 0 & 0 & 0 & 0 & 1 & 0 & & & \\ 0 & 0 & 0 & 0 & 0 & 1 & & & \\ \hline & & & 0 & 0 & 0 & 1 & 0 & 0 \\ & & & 0 & 0 & 0 & 0 & 1 & 0 \\ & & & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline & & & & & & 0 & 0 & 0 \\ & & & & & & 0 & 0 & 0 \\ & & & & & & 0 & 0 & 0 \end{array} \right].$$

Perform the same similarity transformation on the matrices B and C to get the matrices

$$K = D^{-1}BD = \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 3 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline & & & 1 & 0 & 3 & 0 & 3 & 0 \\ & & & 0 & 1 & 0 & 0 & 0 & 0 \\ & & & 0 & 0 & 1 & 0 & 0 & 0 \\ \hline & & & & & & 1 & 0 & 3 \\ & & & & & & 0 & 1 & 0 \\ & & & & & & 0 & 0 & 1 \end{array} \right],$$

$$L = D^{-1}CD = \left[\begin{array}{ccc|ccc|ccc} 4 & 0 & 0 & 0 & 0 & 3 & 3 & -3 & 0 \\ 0 & 1 & 0 & 0 & 0 & 3 & 0 & 0 & 3 \\ 0 & 0 & 4 & 0 & 3 & 0 & 0 & 0 & 0 \\ \hline & & & 4 & 0 & 0 & 0 & 0 & 3 \\ & & & 0 & 1 & 0 & 0 & 0 & 3 \\ & & & 0 & 0 & 4 & 0 & 3 & 0 \\ \hline & & & & & & 4 & 0 & 0 \\ & & & & & & 0 & 1 & 0 \\ & & & & & & 0 & 0 & 4 \end{array} \right].$$

(As a numerical shortcut, note that since the matrices K and L must centralize the Weyr matrix W , we know K and L once we have computed their top row of blocks.) Our problem now is reduced to calculating the dimension of the subalgebra $\mathcal{B} = F[W, K, L]$ generated by W, K, L . (Since \mathcal{A} and \mathcal{B} are isomorphic as algebras under conjugation by D , certainly $\dim \mathcal{A} = \dim \mathcal{B}$.) The advantage of working with \mathcal{B} is that it contains the nilpotent Weyr matrix W of known Weyr structure, and we

can look at its associated leading edge subspaces. The calculations from here on are a breeze.

Let U_0, U_1, U_2 be the leading edge subspaces of \mathcal{B} relative to W . The $(1, 1)$ blocks of W^0 , K and L contribute to U_0 the matrices

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

These three matrices generate a commutative subalgebra of $M_3(F)$ of dimension 3, and one can quickly check that the subalgebra is self-centralizing. (Later in Chapter 5 we will establish a result of Laffey and Lazarus, and Neubauer and Saltman, that says commutative subalgebras of $M_n(F)$ that are 2-generated in the algebra sense and of vector space dimension n are always self-centralizing.) Therefore by Proposition 3.4.4 (1), $\dim U_0 = 3$. By Proposition 3.4.4 (6), since U_0 is a self-centralizing subalgebra of $M_3(F)$ and the Weyr matrix $W \in \mathcal{B}$ has 3 blocks in its homogeneous Weyr structure, we must have

$$\dim \mathcal{A} = \dim \mathcal{B} = \dim U_0 \times 3 = 9.$$

This completes our task. Notice by Proposition 3.4.4 (3), (4) that $U_0 = U_1 = U_2$ in this example. \square

Example 3.5.2

In this example, we examine a family of 3-generated commutative subalgebras of $M_n(F)$ in which the first generator is nilpotent and already in Weyr form. The family¹² will be of interest to us in Chapter 7.

Fix a positive integer s and let $n = 4s$. In terms of blocked matrices whose entries are $s \times s$ matrices, let

$$W = \left[\begin{array}{cc|cc} 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & I \\ & & & 0 \end{array} \right], K = \left[\begin{array}{cc|cc} 0 & A & B & C \\ 0 & 0 & 0 & D \\ \hline & & 0 & B \\ & & & 0 \end{array} \right], K' = \left[\begin{array}{cc|cc} 0 & A' & B' & C' \\ 0 & 0 & 0 & D' \\ \hline & & 0 & B' \\ & & & 0 \end{array} \right].$$

Note that W is a nilpotent Weyr matrix with Weyr structure $(2s, s, s)$. Also K and K' are nilpotent of index at most 3. By Proposition 3.2.1, W commutes with K and K' . The condition that K and K' commute is easily seen to be

$$(*) \quad AD' + BB' = A'D + B'B.$$

12. This family was studied by R. M. Guralnick in 1992, although he didn't use the Weyr form.

Suppose $(*)$ holds and let $\mathcal{A} = F[W, K, K']$ be the commutative subalgebra of $M_n(F)$ generated by W, K, K' . Since all products of any three proper powers of W, K, K' are zero, \mathcal{A} is spanned as a vector space by $I, W, K, K', W^2, K^2, (K')^2, KK', KW, K'W$, whence $\dim \mathcal{A} \leq 10$. So, even for large n , these subalgebras are quite small. Nevertheless, they are of considerable interest to us in Chapter 7 when we establish a result by Guralnick, which turns out to have a direct impact on an approximate simultaneous diagonalization question (which in turn has modern relevance to certain questions in biomathematics and multivariate interpolation). For his result, Guralnick requires $n \geq 32$. But in the present discussion we will choose a small n and be content to see what the leading edge subspaces of \mathcal{A} look like.

Let us set $s = 2$ so that W is an 8×8 Weyr matrix with Weyr structure $(4, 2, 2)$. In the choice of K and K' , set

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, D = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

$$A' = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, B' = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, C' = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, D' = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

One checks that the above condition $(*)$ holds so W, K, K' commute.

Let U_0, U_1, U_2 be the leading edge subspaces of \mathcal{A} relative to W . Since U_0 is generated as an algebra by the top left 4×4 corners of W, K, K' , we see that U_0 has a basis

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

(The identity block comes not from W but from the identity of $M_8(F)$ —remember subalgebras must contain the identity.) Thus, $\dim U_0 = 3$.

What 4×2 matrices can be in U_1 ? Certainly

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \in U_1$$

because this is the leading (nonzero) block of W . In top row notation, suppose $L = [0, X, Y] \in \mathcal{A}$. Since $I, W, K, K', W^2, K^2, (K')^2, KK', KW$, and $K'W$ span \mathcal{A} , the matrix L is a linear combination of these, but the combination can't include I, K, K' because they are linearly independent in the top left 4×4 corner (and the others are zero there). Among the remaining matrices in the linear combination,

only W has a nonzero $(1, 2)$ block. Hence, X is a scalar multiple of $\begin{bmatrix} I \\ 0 \end{bmatrix}$. Thus, $\dim U_1 = 1$. This contrasts with what happens in the homogeneous case (Proposition 3.4.4 (3)), where always $\dim U_1 \geq \dim U_0$.

Next we examine U_2 . Suppose in top row notation, $L = [0, 0, Z] \in \mathcal{A}$. By the same argument as in the previous paragraph, L must be a linear combination of W^2 , K^2 , $(K')^2$, KK' , KW , and $K'W$. These six matrices are of the form $[0, 0, *]$ and so contribute to U_2 . Thus, U_2 is spanned by their $(1, 3)$ blocks, and they in turn are all of the form

$$\begin{bmatrix} * & * \\ * & * \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Therefore, $\dim U_2 \leq 4$. However, W^2 , K^2 , $(K')^2$, KW , respectively, contribute to U_2 the four independent matrices

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

whence $\dim U_2 = 4$.

Finally, we compute the dimension of our commutative subalgebra according to Theorem 3.4.3:

$$\begin{aligned} \dim \mathcal{A} &= \dim U_0 + \dim U_1 + \dim U_2 \\ &= 3 + 1 + 4 \\ &= 8. \end{aligned}$$

□

If one is trying to construct a commutative subalgebra $\mathcal{A} = F[A_1, A_2, \dots, A_k]$ of $M_n(F)$, generated by k commuting matrices A_1, A_2, \dots, A_k , such that $\dim \mathcal{A}$ exceeds a prescribed bound (such as n in the case of 3 generators), then the Weyr form and leading edge subspaces suggest a promising method of attack. We close this chapter with a brief discussion of the method.

Firstly, if such a subalgebra \mathcal{A} (as in the previous paragraph) exists, then there is one with a minimal n and with A_1 a nilpotent Weyr matrix of a certain Weyr structure (n_1, n_2, \dots, n_r) . Projecting \mathcal{A} onto its top left $(r-1) \times (r-1)$ corner of blocks is an algebra homomorphism, under which A_1 remains in Weyr form but with Weyr structure $(n_1, n_2, \dots, n_{r-1})$ (the last block has been removed). The commutative subalgebra generated by

the $(r-1) \times (r-1)$ corners of the A_i cannot, by minimality of n , have the required dimension (relative to matrix size $n - n_r$). But somehow, the desired example of matrix size n arises by the addition of just one more row and column of blocks. That suggests the strategy of constructing the corners of the A_i successively in terms of the corners of A_1 having in turn Weyr structures

$$\begin{aligned} & (n_1) \\ & (n_1, n_2) \\ & (n_1, n_2, n_3) \\ & \vdots \\ & (n_1, n_2, n_3, \dots, n_r). \end{aligned}$$

In going from matrix size $n_1 + n_2 + \dots + n_j$ to the next matrix size $n_1 + n_2 + \dots + n_j + n_{j+1}$, we only have to choose *one* new block for each A_i for $i = 2, \dots, r$, namely the new $(1, j+1)$ block. This is because of the form of the centralizer of a Weyr matrix—remember matrices in the centralizer of A_1 are completely determined by their top rows. Moreover, in checking that the larger A_i still commute, we only have to check that the products $A_p A_q$ and $A_q A_p$ agree in the $(1, j+1)$ block (again because of the form of matrices in the centralizer, and the homomorphic fact that the products already agree in the other first row blocks). At each step, one keeps track of the dimension of the new leading edge subspace U_j and keeps a tally of $\dim U_0, \dim U_1, \dots, \dim U_j$. By Theorem 3.4.3, the sum of these dimensions is the dimension of the $(j+1) \times (j+1)$ corner of our algebra \mathcal{A} . Notice (again using the homomorphic property of the corner projections) that, at each step, the earlier leading edge subspaces U_0, U_1, \dots, U_{j-1} associated with the algebra of corners so far constructed will be unchanged in the new algebra of bigger corners. This is important to remember.

Example 3.5.3

To illustrate the above strategy, suppose we suspect we can achieve $\dim F[A_1, A_2, A_3] > n$ for suitable commuting $n \times n$ matrices A_1, A_2, A_3 with A_1 a nilpotent Weyr matrix of homogeneous Weyr structure $(4, 4, 4, \dots, 4)$. Let's use the notation

d_0	d_1	d_2	\dots	d_{r-1}
-------	-------	-------	---------	-----------

to indicate that the dimensions of the leading edge subspaces $U_0, U_1, U_2, \dots, U_{r-1}$ of an algebra (relative to some understood Weyr structure) are, respectively, $d_0, d_1, d_2, \dots, d_{r-1}$. Since we are trying to make the dimension of our algebra $\mathcal{A} = F[A_1, A_2, A_3]$ large, it is tempting to start our construction by making the

2	4×4 corner
2 4	8×8 corner
2 4 6	12×12 corner

$$\begin{aligned} D_0 &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad E_0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ D_1 &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad E_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ D_2 &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

13. To see that the leading edge subspace U_2 has dimension 6, note the six linearly independent contributions to U_2 from W^2 , A_1W^2 , A_2W , A_1A_2W , $A_1^2 - A_3$, $A_1A_3 - A_2^3$.

done by hand—one was really only faced at each step with multiplying a few 4×4 matrices, despite the full matrices being 16×16 at the fourth step. The Weyr form does make life easier in commuting problems. \square

BIOGRAPHICAL NOTE ON FROBENIUS

Ferdinand Georg Frobenius was born in Berlin on October 26, 1849. He received his doctorate at the University of Berlin in 1870, under the supervision of Weierstrass. In 1892, Frobenius took the mathematics chair at the University of Berlin, and in the following year, he was elected to the Prussian Academy of Sciences. Frobenius is principally known for his work in differential equations and finite group theory, particularly through his contributions to group representations and character theory. But today's undergraduate mathematics student should also be grateful to Frobenius for the first full proof of the Cayley–Hamilton theorem, and the Sylow theorems for abstract groups (as against permutation groups). Frobenius supported the Berlin view that applied mathematics belonged to technical schools, not universities! He is also remembered for a serious underrating of Hilbert's potential, when in a letter of recommendation for the latter for an appointment at Göttingen he wrote: "He is a rather good mathematician, but will never be as good as Schottky." Frobenius died in Berlin on August 3, 1917.

The Module Setting

Modules can provide great insights in algebra. In this chapter, we formulate the Weyr form module-theoretically and show how this leads naturally to our third way of establishing the existence of the Weyr form for matrices. The generality of our setting is quite surprising, as are the relatively simple arguments involved, at least for those familiar with basic ring and module theory. However, we do not assume all our readers have this familiarity, so in keeping with our philosophy of making the book largely self-contained, we develop the necessary module theory from scratch. Since later chapters can be read independently of this one, the reader also has the choice of simply skipping this chapter. For cultural reasons, we hope this option is not exercised.

In Sections 4.1 to 4.4 we introduce the basics of module theory, concentrating on the facets of the theory that are pertinent to our goal. On the other hand, we do assume that the reader has at least a nodding acquaintance with (additive) abelian groups, subgroups, factor groups, rings, (one-sided) ideals, factor rings, homomorphisms, and direct sum decompositions. These concepts are the standard fare of most introductory texts on abstract algebra, for example, Nicholson's *Introduction to Abstract Algebra*, or Jacobson's *Basic Algebra I*.

The central module concept for us is that of a projective module (over an arbitrary noncommutative ring). The key ring concept is the notion of a von

Neumann regular element (particularly in the ring of module endomorphisms of some projective module), which is an element possessing a quasi-inverse. Von Neumann regular rings are rings in which all elements have this property, and they were introduced by von Neumann to co-ordinatize certain lattices. We need very little of the well-developed machinery of regular rings, say as expounded in Goodearl's excellent book *Von Neumann Regular Rings*, but the first lemma in Goodearl's Chapter 7, after a little modification, is crucial to our approach.

There is a long history of ring-theoretic arguments providing striking insights into linear algebra and its applications. Amongst the most successful of these, for readers familiar with the area, is in the study of group representations of a finite group G over an algebraically closed field F . Here a *representation of degree n* is simply a group homomorphism of G into the group $GL_n(F)$ of $n \times n$ nonsingular (invertible) matrices over F , or equivalently, a group homomorphism of G into the group $GL(V)$ of invertible linear transformations of an n -dimensional vector space V over F . The study of group representations (and its associated character theory) has proved an indispensable tool for revealing the structure of finite groups. The contribution of ring theory came with the realization that one can form the so-called group algebra $F[G]$, a finite-dimensional associative algebra over F having the group elements as a basis and whose multiplication extends the group multiplication. Then the representations of G correspond exactly to the finite-dimensional modules over the ring $F[G]$. In the case where the characteristic of F does not divide the order of G , Maschke's theorem describes the ring structure of $F[G]$ as a finite direct product of various full matrix algebras $M_{n_i}(F)$. Then the $F[G]$ -module structures are known, whence too are the representations of G (in particular, the various n_i are the degrees of the "irreducible" representations of G , and the sum of their squares is the order of G). There is so much more to this story, which we will not pursue in this book (the reader can refer to Chapter 5 of Jacobson's *Basic Algebra II*). But the point we wish to make is that while group representations can be studied without using this ring and module structure (and they frequently are, by physicists and chemists), the additional insight gained by doing so is most rewarding.

The Jordan form can be established by module-theoretic arguments, namely, using the known structure of finitely generated modules M over a principal ideal domain R , as direct sums of cyclic modules. This is a favorite application of module theory and is covered in many texts. In the case of the Jordan form of an $n \times n$ matrix A over an algebraically closed field F , the relevant module M is the space of column vectors F^n , the relevant ring R is the polynomial ring $F[x]$, and the module action is through polynomials in A multiplying column vectors. In Section 4.6, we will derive the Jordan form this way but without developing the

necessary module structure theory, which would take us too far afield. We can then use this to contrast our later derivation of the Weyr form in Section 4.8, in three important respects. First, for the modules M we use in the Weyr form, there is no restriction on the ring R and only the projective (in fact, only quasi-projective) restriction on M (no finitely generated restriction). We decompose M relative to a given nilpotent endomorphism, all of whose powers are regular. Second, compared with the work required to establish the theorem on finitely generated modules over principal ideal domains, the structure decomposition required of our M is easier to derive. Third, getting the Weyr form of a matrix as a corollary this way suggests that the Weyr form is more “basis-free” than its Jordan counterpart, that is, its description need not reference a basis or 1-dimensional subspaces. In short, the Weyr form seems to live in a somewhat bigger universe and is perhaps more natural. We expand on this comparison in Section 4.9.

As we discussed in Chapter 1, canonical forms quickly reduce to the case of a nilpotent matrix. Our derivation of the Weyr form in the nilpotent case picks on a feature that one initially tends to dismiss as not being that critical, namely, the powers of the matrix are (von Neumann) regular. After all, all matrices are regular. Indeed, even all transformations of an infinite-dimensional vector space are regular. However, it turns out, as we show in Section 4.10, insisting that the powers of a nilpotent element a in an *arbitrary* ring A be regular has the surprising consequence that, “locally,” a sits inside A much like a matrix in Jordan form (or a matrix in Weyr form). For instance, if A happens to be an algebra over a field F , there exists an element b in A and an isomorphism of the subalgebra $F[a, b]$ generated by a and b into some matrix algebra $M_n(F)$ such that, under the isomorphism, a is in Jordan (or Weyr) form, b is its transpose, and $F[a, b]$ is the subalgebra of all block diagonal matrices in $M_n(F)$ having the same block structure as a . There is a more general statement when A is an algebra over any commutative ring Λ . As we demonstrate in Section 4.11, it is crucial that *all* powers of the nilpotent element a be regular in order to reach our conclusions.

4.1 A MODICUM OF MODULES

The concept of a module M over a given ring R is a powerful tool of ring theory, whose potential was first recognized by Emmy Noether in the late 1920s. It generalizes a vector space over a field. In this section we lay down the very basics. The next two sections then develop some core material of modules needed later for our particular applications to the Weyr form.

We assume our ring R has a multiplicative identity 1, but R need not be a commutative ring. If one looks at the axioms for a vector space M over a field F ,

one notices that they still make sense if the scalars come from a general ring R . It is a little glib to say that such structures constitute modules M over R , since there are actually two types of modules, *left* and *right*. We concentrate on the former.¹ Here is the formal definition.

Definition 4.1.1: Let R be a ring. A **left R -module** is an abelian group M (written additively) together with a multiplication $r \cdot x \in M$ of members x of M by members r of R such that for all $x, y \in M$ and $r, s \in R$:

- (1) $r \cdot (x + y) = r \cdot x + r \cdot y$
- (2) $(r + s) \cdot x = r \cdot x + s \cdot x$
- (3) $(rs) \cdot x = r \cdot (s \cdot x)$
- (4) $1 \cdot x = x$

Just as with vector spaces, we usually omit the ‘ \cdot ’ in $r \cdot x$ and write rx . The associative axiom (3) then reads $(rs)x = r(sx)$ and involves both the ring multiplication and the module multiplication. But the context makes clear where the multiplication is taking place. Also, we loosely refer to M itself as the module if the action of R in the product rx is understood.

In performing “arithmetic” within a given left R -module M , one should feel just as comfortable as with the corresponding arithmetic of a vector space (using properties such as $r \cdot 0 = 0$, $0 \cdot x = 0$, $(-r) \cdot x = -(r \cdot x) = r \cdot (-x)$), but with three important provisos: $rx = 0$ does not necessitate $x = 0$ (the zero of the abelian group M) or $r = 0$ (the zero of the ring R), (2) $rx = ry$ and $r \neq 0$ does not imply $x = y$, and (3) $rx = sx$ and $x \neq 0$ does not imply $r = s$. See Example 4.1.3 below.

The qualifying *left* in a left R -module indicates that the elements of M are multiplied on the left by ring elements. A similar definition gives the corresponding notion of a **right R -module**, with the elements x of M multiplied on the right by elements r of R to give xr . If we are lucky enough to have R commutative, then every left R -module becomes a right R -module by letting $x \cdot r = rx$. However, this left–right transfer fails for general rings, the right-hand version of property (3) being the stumbling block.

For the sake of brevity, from now on we’ll use the unqualified term *module* to mean *left module*.

1. The authors agonized over this. Given our convention for expressing function values $f(x)$, and composing functions $(f \circ g)(x) = f(g(x))$, it makes more sense to work primarily with right modules, to avoid certain “anti-isomorphic twists.” What swayed us to the left was the belief that our reader would more than likely write scalars on the *left* of vectors in linear algebra. And we didn’t feel like embarking on a crusade to change that.

Example 4.1.2

When R is a field, the R -modules are exactly the vector spaces over R . \square

Example 4.1.3

Let $R = \mathbb{Z}$ be the ring of integers. Then *any* abelian group $(M, +)$ becomes an R -module upon defining $n \cdot x$ to be the n th multiple nx of x :

$$nx = \begin{cases} x + x + \cdots + x, & \text{the sum of } n \text{ copies of } x \text{ if } n > 0, \\ -x - x - \cdots - x, & \text{the sum of } -n \text{ copies of } -x \text{ if } n < 0, \\ 0 & \text{if } n = 0. \end{cases}$$

Moreover, every \mathbb{Z} -module M must take this form. For instance, when $n > 0$, by axioms (2) and (4) we have

$$\begin{aligned} n \cdot x &= (1 + 1 + \cdots + 1) \cdot x \\ &= 1 \cdot x + 1 \cdot x + \cdots + 1 \cdot x \\ &= x + x + \cdots + x. \end{aligned}$$

Notice that when M is not a *torsion-free* group, that is, M has a nonzero element of finite order, we can have $rx = 0$ with neither x nor r being zero. (Take x of order $r > 0$.)

On the other hand, not every abelian group $(M, +)$ can be converted into a module over the ring of rational numbers \mathbb{Q} . Readers may wish to convince themselves that a necessary and sufficient condition for this is that M be a torsion-free group that is also a *divisible* group, that is for any $x \in M$ and positive integer n , there exists $y \in M$ with $x = ny$. \square

Example 4.1.4

Any ring R can be regarded as a left module over itself by taking the ring multiplication as the module multiplication: $r \cdot x = rx$, namely the ring product of r and x , for all $r \in R$, $x \in R$. (The addition $+$ in the module is of course the ring addition.) This module turns out to be important for describing a general R -module (see Section 4.3). \square

Example 4.1.5

For positive integers m , n , and a ring R , let $M_{m \times n}(R)$ denote the additive group of $m \times n$ matrices with entries from R . Denoting such a matrix as usual by (r_{ij}) , we see that $M_{m \times n}(R)$ becomes an R -module if we define $r \cdot (r_{ij}) = (rr_{ij})$ for all $r \in R$. \square

We began by motivating a module over a ring R as a natural generalization of a vector space. To a ring theorist, that is only half the story. The second half

concerns representations² of R , and this point of view is not really present in the study of vector spaces over a fixed field F . We won't make much use of the connection in our book, but nevertheless it is useful to be at least aware of the connection, if for no other reason than it often helps one recognize when a nice module is sitting in the background, waiting to be asked to dance. Motivation of the representation connection can be done through an analogue of Cayley's theorem for groups. Recall the latter says that every group G is (isomorphic to) a subgroup of the symmetric group S_X on some set X . The proof is really easy.³ Simply let $X = G$ and define a 1-1 group homomorphism $\theta : G \rightarrow S_X$ by letting $\theta(g)$ be the permutation $x \mapsto gx$. The analogue of Cayley's theorem for a ring R is that for each $r \in R$, the map $\theta_r : R \rightarrow R, x \mapsto rx$ is an endomorphism⁴ of the additive group $(R, +)$, and then the map $\theta : R \rightarrow R, r \mapsto \theta_r$ is a 1-1 ring homomorphism of R into the ring $\text{End}(R)$ of all group endomorphisms of $(R, +)$. Here, we define the addition and multiplication of endomorphisms f and g by

$$\begin{aligned}(f + g)(x) &= f(x) + g(x) \\ fg &= f \circ g,\end{aligned}$$

where $f \circ g$ is our agreed way of composing two functions (g first, followed by f). But now this suggests looking more generally at (not necessarily 1-1) ring homomorphisms $\theta : R \rightarrow \text{End}(M)$ from R into the ring $\text{End}(M)$ of all group endomorphisms of some abelian group $(M, +)$. Such a representation is really just a left R -module. We record this in our next family of examples. An important theme in ring theory is that “nice rings have nice modules,” and one can often obtain structural information about the ring R from a knowledge of its modules. However, we shall not pursue this line in our book.

2. In an algebraic context, a “representation” refers to a homomorphism of an abstract algebraic structure, such as a group or an associative algebra, into a more concrete structure of the same type, whose members may be permutations or matrices, for example. One can then often make use of additional “arithmetic functions” associated with the concrete objects, such as the cycle structure of a permutation or the trace of a matrix.

3. For this reason, the importance of Cayley's theorem is often underestimated. The first author once heard it described (by a colleague who should have known better) as “the most useless theorem in algebra.” The importance of Cayley's theorem lies not so much in this particular “regular representation,” but in the fact that it suggests looking at other permutation representations, and more generally representations of other algebraic structures, particularly through the use of matrices. And that really *is* important.

4. A homomorphism of an algebraic structure into itself is called an *endomorphism*.

Example 4.1.6

Let R be a ring and let $\theta : R \rightarrow \text{End}(M)$ be a ring homomorphism from R into the ring $\text{End}(M)$ of all group endomorphisms of some abelian group $(M, +)$. Then M becomes a left R -module under the action

$$r \cdot x = \theta(r)(x) \text{ for all } x \in M \text{ and } r \in R.$$

The four defining properties for a module now correspond respectively to (1) $\theta(r)$ is a group endomorphism, (2) the definition of the sum of two endomorphisms, (3) the definition of the product of two endomorphisms, and (4) our ring homomorphisms preserving the identity. An important special case of this construction of a module is when R already sits inside $\text{End}(M)$ as a subring and θ is the inclusion map. For instance, if R is any subring of the ring of linear transformations of some vector space M , we can give M the structure of an R -module by taking $r \cdot x = r(x)$, namely the value of r at x .

Conversely, every left R -module can be viewed as arising from some ring homomorphism $\theta : R \rightarrow \text{End}(M)$ for a suitable abelian group $(M, +)$. To check this statement, given the module M , use its additive structure for the abelian group and define $\theta(r) : M \rightarrow M$ by $\theta(r)(x) = r \cdot x$. \square

The following is a very important example of a module, which we use in Section 4.6 to derive the Jordan form of a matrix A . It is also a good illustration of how an understanding of the above connection between modules and representations can suggest a natural module for the task at hand.

Example 4.1.7

Let F be a field and V a vector space over F . Let $t : V \rightarrow V$ be a fixed linear transformation. We can use t to make V into a module over the polynomial ring $R = F[x]$ as follows. The substitution map

$$\theta : F[x] \rightarrow \mathcal{L}(V), f(x) \mapsto f(t)$$

is a ring homomorphism from R into the ring $\mathcal{L}(V)$ of linear transformations of V . Inasmuch as $\mathcal{L}(V)$ is a subring of the ring $\text{End}(V)$ of endomorphisms of the abelian group V , by Example 4.1.6 we have a module action

$$f(x) \cdot v = \theta(f)(v) = f(t)(v) \text{ for all } v \in V \text{ and } f(x) \in F[x].$$

In other words, a polynomial $f(x)$ acts on a vector $v \in V$ by evaluating the transformation $f(t)$ at v .

A special case of this, and the one of interest when deriving the Jordan form in Section 4.6, is to start with a matrix $A \in M_n(F)$, let V be the space F^n of all $n \times 1$ column vectors over F , and let $t : V \rightarrow V$ be the linear transformation that left

multiplies column vectors by the matrix A . Then the module action of $R = F[x]$ on V is simply

$$f(x) \cdot v = f(A)v \text{ for all } v \in V \text{ and } f(x) \in F[x].$$

The product $f(A)v$ is the matrix product of the $n \times n$ matrix $f(A)$ and the $n \times 1$ matrix v . For instance, let

$$A = \begin{bmatrix} 1 & 1 & -2 \\ 0 & -1 & 3 \\ -3 & 1 & 1 \end{bmatrix}.$$

Then $V = F^3$ and, for example, the polynomial $f(x) = 1 + x + x^2$ acts on the vector

$$v = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$$

through the recipe

$$\begin{aligned} f(x) \cdot v &= (I + A + A^2)v \\ &= \begin{bmatrix} 9 & -1 & -3 \\ -9 & 4 & 3 \\ -9 & -2 & 12 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 28 \end{bmatrix}. \quad \square \end{aligned}$$

This is a good place to make precise the notion of an *algebra over a commutative ring*. It is a term we have informally used several times in earlier chapters, beginning in Section 1.1 of Chapter 1 when we talked about $M_n(F)$ as the “algebra” of $n \times n$ matrices over the field F . The formal definition that follows integrates three concepts—a given ring A (with identity but not necessarily commutative), a commutative ring Λ , and a module action of Λ on A that intertwines with the ring product in A . In the matrix algebra example, $A = M_n(F)$, $\Lambda = F$, and the module action is the usual scalar multiplication of matrices.

Definition 4.1.8: Let A be a ring with identity and Λ a commutative ring (also with an identity). Then A is an **algebra over Λ** if A is a left Λ -module relative to a module action ‘ \cdot ’ that satisfies

$$\lambda \cdot (ab) = (\lambda \cdot a)b = a(\lambda \cdot b)$$

for all $\lambda \in \Lambda$ and $a, b \in A$.

We normally suppress the ‘ \cdot ’ in the module action, so the intertwining relation becomes

$$(*) \quad \lambda(ab) = (\lambda a)b = a(\lambda b).$$

The multiplications are unambiguous provided we know that λ comes from Λ and a, b from A . We also use the same symbol 1 to denote the identities of A and Λ . What $(*)$ amounts to is saying the map $\theta : \lambda \mapsto \lambda \cdot 1$ is a ring homomorphism of Λ into the center $\mathcal{C}(A)$ of A such that $\lambda \cdot a = \theta(\lambda)a$ (and where the product on the right is the ring product in A). This also provides a good way of seeing if a given ring A will support an algebra structure over a given commutative ring Λ —does A contain a homomorphic image of Λ in its center? (In the case of a field Λ , the image must be an isomorphic copy.) If so, and $\theta : \Lambda \rightarrow \mathcal{C}(A)$ is a ring homomorphism (preserving the identity), then A becomes an algebra over Λ under the module action $\lambda \cdot a = \theta(\lambda)a$. Another good way of looking at $(*)$ is that the left and right multiplication maps of A by fixed members $a, b \in A$ are now Λ -endomorphisms of the left Λ -module A .

Every ring A can be regarded as an algebra over the ring \mathbb{Z} of integers, where \mathbb{Z} acts on A just as an additive abelian group (Example 4.1.3). Outside of the present chapter, all our algebras A are over a field F . Thus, $M_n(F)$ is a good example. The ring $\mathbb{H} = \mathbb{R}[i, j, k]$ of real quaternions is also a good example of an algebra over the real field \mathbb{R} . Since \mathbb{H} contains a copy of the complex field \mathbb{C} , namely

$$C = \{a + bi : a, b \in \mathbb{R}\} \subseteq \mathbb{H},$$

the ring \mathbb{H} is naturally a 2-dimensional vector space over \mathbb{C} . A common mistake is to assume that this will give \mathbb{H} the structure of an algebra over \mathbb{C} . Wrong: C does not lie inside the center of \mathbb{H} . (Besides, since \mathbb{C} is an algebraically closed field, the only finite-dimensional complex algebra without divisors of zero is \mathbb{C} itself.) But we are getting off the track a little here. Let's return to modules.

The notions of *submodule* and *factor module* of a module M over R are the obvious extensions of their vector space counterparts: A **submodule** of M is a subgroup N of $(M, +)$ that is also closed under the module multiplication by R (thus, N becomes an R -module itself under the restriction of the multiplication by R to N). The **factor module** M/N is then the factor group $(M/N, +)$ under the module action

$$r \cdot (x + N) = rx + N.$$

The singleton set $\{0\}$ consisting of the zero of M (which in the future we denote simply by 0) and the full module M are always submodules. Here are the

submodules of some of the modules we discussed earlier:

Module	Submodules
Vector space M (Example 4.1.2)	subspaces of M
\mathbb{Z} -module M (Example 4.1.3)	subgroups of M
Ring R as left R -module (Example 4.1.4)	left ideals of R
F^n as $F[x]$ -module (Example 4.1.7)	t -invariant subspaces of F^n

Given two left modules M and N over the same ring R , an **R -homomorphism** from M into N is just the module analogue of a linear transformation: it is a mapping $f : M \rightarrow N$ satisfying

- (i) $f(x + y) = f(x) + f(y)$
- (ii) $f(rx) = rf(x)$

for all $x, y \in M$, and $r \in R$. A **module isomorphism** of course is just a homomorphism that is also a bijection, and we write $M \cong N$ if there exists a module isomorphism from M onto N . As expected, the fundamental theorem for homomorphisms still applies.

Theorem 4.1.9 (Fundamental Homomorphism Theorem)

If we have an R -homomorphism $f : M \rightarrow N$, then the kernel $\ker(f) = \{x \in M : f(x) = 0\}$ is a submodule of M , the image $\operatorname{im}(f) = \{f(x) : x \in M\}$ is a submodule of N , and

$$\operatorname{im}(f) \cong M / \ker(f).$$

Also as expected, an onto homomorphism $f : M \rightarrow N$ is an isomorphism precisely when $\ker(f) = 0$, and then its inverse function $f^{-1} : N \rightarrow M$ is also an R -homomorphism.

An R -homomorphism of a module M to itself is called an **R -endomorphism** of M . The set of all such endomorphisms is denoted by $\operatorname{End}_R(M)$ and is a ring under pointwise addition and function composition: $(f + g)(x) = f(x) + g(x)$, $(fg)(x) = f(g(x))$.

Module-theorists love arrow diagrams, believing that the eye is quicker than the hand when it comes to composing homomorphisms. To them, a good “diagram chase” is as exciting as a horse race.⁵ Here a homomorphism $g : M \rightarrow N$ is represented by an arrow going from M to N , labeled by g , and depicted horizontally, vertically, or diagonally, as befits the situation.

5. Some have been known to don a riding hat and riding boots for a tight chase.

The diagram is said to be **commutative** if the composition of the module homomorphisms along any two directed paths from one module in the diagram to another yield the same homomorphism. In these terms, our definition of the composition $f \circ g$ of two homomorphisms $g : M \rightarrow N$ and $f : N \rightarrow P$ amounts to saying the diagram

$$\begin{array}{ccc} & M & \\ f \circ g \swarrow & \downarrow g & \\ P & \xleftarrow{f} & N \end{array}$$

is commutative.

Given a (left) R -module M and a subset $S \subseteq M$, the **annihilator** of S in R is

$$\text{ann}_R(S) = \{r \in R : rs = 0 \text{ for all } s \in S\}.$$

This is a left ideal of R and is a two-sided ideal if S is an R -submodule. For a submodule N of M , if $I = \text{ann}_R(N)$, we can give N a left module structure over the factor ring R/I by letting $r + I \in R/I$ act on $x \in N$ by

$$(r + I) \cdot x = rx$$

where rx is the product in the module M .

A **cyclic submodule** takes the form $Ra = \{ra : r \in R\}$ for some $a \in M$. Thus, for vector spaces, the nonzero cyclic submodules are the 1-dimensional subspaces, while for abelian groups, the cyclic \mathbb{Z} -submodules are its cyclic subgroups. Notice that cyclic modules of a ring R are directly related to the structure of the ring. They are (to within isomorphism) just the factor modules R/I as I ranges over the left ideals of R . (If $M = Ra$ is a cyclic module, just apply the fundamental homomorphism theorem to the mapping $R \rightarrow M$, $r \mapsto ra$.)

The **sum** $M_1 + M_2 + \cdots + M_k$ of two or more submodules is defined exactly as for subspaces. In particular, the smallest submodule of a module M over R that contains specified elements a_1, a_2, \dots, a_k is $Ra_1 + Ra_2 + \cdots + Ra_k$, and is called the **submodule generated by** a_1, a_2, \dots, a_k . A submodule N is said to be **finitely generated** if N is of the form $Ra_1 + Ra_2 + \cdots + Ra_k$ for some $a_1, a_2, \dots, a_k \in M$. As we will see in Section 4.6, the key to understanding the Jordan form of a matrix $A \in M_n(F)$ from a module standpoint is the way the $F[x]$ -module F^n discussed in Example 4.1.7 breaks up as a direct sum of cyclic modules. But we are getting ahead of ourselves here.

We close this section with one other important concept. A **simple** (or **irreducible**) **module** is a nonzero module M that has $\{0\}$ and M as its only submodules. More generally, a **simple submodule** of M is a submodule N that is simple as a module, that is, it is nonzero and the only nonzero submodule of M contained in N is N itself. For vector spaces, the simple submodules are the 1-dimensional subspaces. The simple \mathbb{Z} -submodules of an abelian group are its subgroups of prime order. Regarding a ring R as a left module over itself, its simple submodules are its minimal left ideals. For instance, for a matrix ring $R = M_n(F)$ over a field F , a typical simple submodule of R is the left ideal consisting of all matrices whose only possible nonzero column is the first. On the other hand, the ring \mathbb{Z} of integers has no minimal ideals (because its ideals are principal and $\mathbb{Z}a$ strictly contains $\mathbb{Z}(2a)$ when $a \neq 0$). Again, in general, the simple R -modules of a ring R relate directly to the ring structure of R , being (isomorphic to) the factor modules R/I as I ranges over the maximal left ideals of R .

4.2 DIRECT SUM DECOMPOSITIONS

A major strategy in the theory of modules is to try to chop a module M up into smaller submodules, in the hope that these submodule offspring have no interaction with each other and are more easily described than their parent M . The decompositions we are talking about are direct sum decompositions, which the reader may already be familiar with in, say, vector spaces or abelian groups. In this section we present the module decomposition generalizations of these that are required in later sections. It is fair to warn readers, however, that we do expect some facility with these tools, not just a knowledge of the definitions.

As with direct sums of other structures, there are two types of direct sums of modules, *external* and *internal*. They do, however, have a close relationship. To keep our discussions simple, we will work with only finite direct sums. Then the **external direct sum** $M_1 \oplus M_2 \oplus \cdots \oplus M_k$ of R -modules M_1, M_2, \dots, M_k is simply the cartesian product $M_1 \times M_2 \times \cdots \times M_k$ of the sets M_1, M_2, \dots, M_k endowed with pointwise module operations:

$$\begin{aligned}(x_1, x_2, \dots, x_k) + (y_1, y_2, \dots, y_k) &= (x_1 + y_1, x_2 + y_2, \dots, x_k + y_k) \\ r \cdot (x_1, x_2, \dots, x_k) &= (rx_1, rx_2, \dots, rx_k)\end{aligned}$$

for all $x_i, y_i \in M_i$ and $r \in R$. Then $M_1 \oplus M_2 \oplus \cdots \oplus M_k$ is clearly an R -module. Our knowledge of its structure is pretty much as good as our knowledge of the individual summands.

Definition 4.2.1: Let R be a ring and M an R -module. We say that M is an **internal direct sum** of submodules M_1, M_2, \dots, M_k if every element $x \in M$ can be written uniquely as

$$x = x_1 + x_2 + \cdots + x_k,$$

where $x_i \in M_i$ for $i = 1, 2, \dots, k$. In this case we write

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_k,$$

and refer to x_i as the ***ith* component** of the element x in this decomposition.

When M is an internal direct sum of M_1, \dots, M_k , clearly M is isomorphic to the external direct sum $M_1 \oplus M_2 \oplus \cdots \oplus M_k$ of M_1, \dots, M_k regarded as R -modules in their own right: the mapping $(x_1, x_2, \dots, x_k) \mapsto x_1 + x_2 + \cdots + x_k$ provides an isomorphism from the external to the internal. Conversely, if M is isomorphic to an external direct sum, there is a matching internal direct sum. In future, we will usually drop the qualifying *internal* or *external* because the context should make it clear which of the two applies. *But as a general rule, we will nearly always work with internal direct sums.*

In practice, just as with direct sums of subspaces or subgroups, checking that a sum is a direct sum is best done through an intersection condition: given submodules M_1, M_2, \dots, M_k of M , we have $M = M_1 \oplus M_2 \oplus \cdots \oplus M_k$ if and only if

- (1) $M = M_1 + M_2 + \cdots + M_k$, and
- (2) $M_i \cap (M_1 + \cdots + M_{i-1} + M_{i+1} + \cdots + M_k) = 0$ for $i = 1, 2, \dots, k$.

Notice that for two submodules, the test for $M = M_1 \oplus M_2$ is simply that $M = M_1 + M_2$ and $M_1 \cap M_2 = 0$. Repeated use of this twosome approach allows one to check for directness $M = M_1 \oplus M_2 \oplus \cdots \oplus M_k$ of arbitrary sums through the following “triangular” conditions:⁶

$$\begin{aligned} M_1 &= M_1 \\ M_1 + M_2 &= M_1 \oplus M_2 \\ (M_1 + M_2) + M_3 &= (M_1 + M_2) \oplus M_3 \end{aligned}$$

6. The reasons for including the first equation are twofold: (1) to provide an apex for the triangle, and (2) to exhibit at least one transparently true statement in the hope of encouraging a flagging reader to continue.

$$\begin{aligned}
(M_1 + M_2 + M_3) + M_4 &= (M_1 + M_2 + M_3) \oplus M_4 \\
&\vdots \\
(M_1 + M_2 + \cdots + M_{k-1}) + M_k &= (M_1 + M_2 + \cdots + M_{k-1}) \oplus M_k
\end{aligned}$$

For instance, the fourth equation is equivalent to $(M_1 + M_2 + M_3) \cap M_4 = 0$, but the real import of this (in combination with the first three equations) is that the submodule $M_1 + M_2 + M_3 + M_4$ of M which we have built up so far is a direct sum $M_1 \oplus M_2 \oplus M_3 \oplus M_4$. This step-by-step buildup is often the way experienced practitioners establish that a module is a direct sum of submodules. Two other points, more or less implicit in these types of arguments, are that if we have a direct sum decomposition

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_k$$

then any regrouping of the summands also gives a direct sum decomposition of M into the respective new summands. And if each M_i decomposes as a direct sum

$$M_i = M_{i1} \oplus M_{i2} \oplus \cdots \oplus M_{i_{n_i}},$$

then M is a direct sum of all the broken-down bits:

$$M = M_{11} \oplus M_{12} \oplus \cdots \oplus M_{1_{n_1}} \oplus M_{21} \oplus M_{22} \oplus \cdots \oplus M_{2_{n_2}} \oplus \cdots$$

Apart from involving some awkward notation (as just witnessed), there is nothing complicated in checking these claims, but our advice to the reader is to skip this exercise unless one has nothing better to do.

It is time for a couple of simple examples.

Example 4.2.2

For a positive integer n , let $\mathbb{Z}/(n)$ be the additive group of integers modulo n , regarded as a module over the ring \mathbb{Z} of integers. (Abstractly, this is just the cyclic group of order n .) Let us decompose $\mathbb{Z}/(60)$ as a direct sum of cyclic groups of prime power order. (Those familiar with the primary decomposition theorem for abelian groups will know there is only one way of doing this, and the orders of the cyclic subgroups involved correspond to the prime power factorization of 60.) Denoting the congruence class of an integer a by \bar{a} , we have the \mathbb{Z} -submodules

$$\begin{aligned}
M_1 &= \langle \bar{15} \rangle = \{\bar{0}, \bar{15}, \bar{30}, \bar{45}\} \\
M_2 &= \langle \bar{20} \rangle = \{\bar{0}, \bar{20}, \bar{40}\}
\end{aligned}$$

of orders 4 and 3, respectively, whence $M_1 \cap M_2 = 0$ by Lagrange's theorem. Hence,

$$M_1 + M_2 = M_1 \oplus M_2$$

and this sum has order $4 \times 3 = 12$. Let

$$M_3 = \langle \overline{12} \rangle = \{\overline{0}, \overline{12}, \overline{24}, \overline{36}, \overline{48}\}.$$

Again by Lagrange's theorem, $(M_1 + M_2) \cap M_3 = 0$ so

$$M_1 + M_2 + M_3 = M_1 \oplus M_2 \oplus M_3.$$

This sum has order $12 \times 5 = 60$ and so must be all of $\mathbb{Z}/(60)$. Therefore, we have

$$\begin{aligned} \mathbb{Z}/(60) &= \langle \overline{15} \rangle \oplus \langle \overline{20} \rangle \oplus \langle \overline{12} \rangle \\ &\cong \mathbb{Z}/(4) \oplus \mathbb{Z}/(3) \oplus \mathbb{Z}/(5), \end{aligned}$$

where the second direct sum is external. By way of example, the unique way of writing $\overline{41}$ as a sum $x_1 + x_2 + x_3$ with each $x_i \in M_i$ is $\overline{41} = \overline{45} + \overline{20} + \overline{36}$. \square

Example 4.2.3

Let F be a field and let $R = M_4(F)$. Regard R as a left module over itself, in which case a direct sum decomposition involves left ideals. Let us decompose R as a direct sum of minimal left ideals (which are the simple submodules of our R -module). That is easy. Let

$$\begin{aligned} I_1 &= \left\{ \begin{bmatrix} * & 0 & 0 & 0 \\ * & 0 & 0 & 0 \\ * & 0 & 0 & 0 \\ * & 0 & 0 & 0 \end{bmatrix} \right\}, & I_2 &= \left\{ \begin{bmatrix} 0 & * & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & * & 0 & 0 \end{bmatrix} \right\} \\ I_3 &= \left\{ \begin{bmatrix} 0 & 0 & * & 0 \\ 0 & 0 & * & 0 \\ 0 & 0 & * & 0 \\ 0 & 0 & * & 0 \end{bmatrix} \right\}, & I_4 &= \left\{ \begin{bmatrix} 0 & 0 & 0 & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & * \end{bmatrix} \right\} \end{aligned}$$

where the $*$ entries are arbitrary. The I_i are clearly left ideals. And here is a case where one can see "by eye" that each $A \in R$ is uniquely a sum of matrices from

I_1, I_2, I_3, I_4 , namely its I_i component is just the matrix obtained from A by setting all columns to zero except the i th. Thus,

$$R = I_1 \oplus I_2 \oplus I_3 \oplus I_4.$$

We leave it as an exercise to show that each I_i is a minimal left ideal. A similar decomposition into “single columns” applies to larger matrix rings $R = M_n(F)$ over a field F , to yield a decomposition into minimal left ideals:

$$R = Re_{11} \oplus Re_{22} \oplus \cdots \oplus Re_{nn}$$

where we have used matrix unit notation for the generators (e_{ij} is the matrix with a 1 in the (i, j) position and 0's elsewhere). These left ideals are isomorphic⁷ as left R -modules because the mapping

$$Re_{ii} \rightarrow Re_{jj}, x \mapsto xe_{ij}$$

is an R -isomorphism (whose inverse is the right multiplication by e_{ji}). Finally, we remark that there are many other ways of decomposing R as a direct sum of minimal left ideals, although the left ideals involved are all isomorphic as modules. For instance, when $n = 2$,

$$R = Re \oplus R(1 - e), \text{ where } e = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \text{ (an idempotent).}$$

See Proposition 4.2.4 below. □

As foreshadowed in the previous example, there is a close connection between direct sum decompositions of a ring R into left ideals and orthogonal idempotents of the ring, which we record in our next proposition. An **idempotent** of R is an element e satisfying $e^2 = e$. Two idempotents e and f are **orthogonal** if $ef = 0 = fe$, and a family of idempotents is said to be **orthogonal** if any two distinct members are orthogonal. A simple, but exceedingly useful, observation is that if e is an idempotent, then so is $1 - e$, and it is orthogonal to e . Moreover, their sum is 1, whence $R = Re \oplus R(1 - e)$ as part of the next proposition.

7. Rings that decompose as a direct sum of isomorphic minimal left ideals are very special. They must be isomorphic to an $n \times n$ matrix ring $M_n(D)$ over a division ring D (noncommutative field). (This can be viewed as a special case of the celebrated Wedderburn–Artin theorem. See Jacobson’s *Basic Algebra II*, Chapter 4.) In particular, the only finite-dimensional algebras over an algebraically closed field F that share this property are the $M_n(F)$.

Proposition 4.2.4

Let R be a ring.

- (1) If $R = I_1 \oplus I_2 \oplus \cdots \oplus I_k$ is a direct sum decomposition of R into left ideals, then there are pairwise orthogonal idempotents e_1, e_2, \dots, e_k of R such that

$$1 = e_1 + e_2 + \cdots + e_k$$

and $I_i = Re_i$ for $i = 1, 2, \dots, k$.

- (2) Conversely, given pairwise orthogonal idempotents e_1, e_2, \dots, e_k of R with $1 = e_1 + e_2 + \cdots + e_k$, there is a direct sum decomposition

$$R = Re_1 \oplus Re_2 \oplus \cdots \oplus Re_k.$$

Proof

(1) Inasmuch as $R = I_1 + I_2 + \cdots + I_k$, we can write $1 = e_1 + e_2 + \cdots + e_k$ for some $e_i \in I_i$. Fix i with $1 \leq i \leq k$. Now

$$\begin{aligned} e_i &= e_i 1 \\ &= e_i(e_1 + e_2 + \cdots + e_k) \\ &= e_i e_1 + e_i e_2 + \cdots + e_i e_k \end{aligned}$$

whence

$$0 = e_i e_1 + \cdots + (e_i^2 - e_i) + \cdots + e_i e_k.$$

But the unique expression of 0 in the direct sum $R = I_1 \oplus I_2 \oplus \cdots \oplus I_k$ is $0 = 0 + 0 + \cdots + 0$. Therefore, we must have $e_i^2 = e_i$ and $e_i e_j = 0$ for $i \neq j$. Thus, e_1, e_2, \dots, e_k are orthogonal idempotents summing to 1. Clearly, $Re_i \subseteq I_i$ because I_i is a left ideal. The reverse containment also holds because for $x \in I_i$ we have $x = x1 = xe_1 + xe_2 + \cdots + xe_k$, which implies $x - xe_i = 0$ by the same argument above that showed $e_i^2 - e_i = 0$. Thus, $x = xe_i \in Re_i$. Hence, $I_i = Re_i$ for $i = 1, 2, \dots, k$.

(2) For any $x \in R$ we have $x = xe_1 + xe_2 + \cdots + xe_k \in Re_1 + Re_2 + \cdots + Re_k$ and therefore $R = Re_1 + Re_2 + \cdots + Re_k$. To establish the directness of this sum, it is enough to show that if $x_i \in Re_i$ for $i = 1, 2, \dots, k$ with $x_1 + x_2 + \cdots + x_k = 0$, then $x_i = 0$ for all i . Fix i . Note that $x_i = r_i e_i$ for some $r_i \in R$. Therefore, since e_1, e_2, \dots, e_k are orthogonal idempotents, we have $x_i e_i = r_i e_i^2 = r_i e_i = x_i$, and $x_i e_j = r_i e_i e_j = 0$ for $i \neq j$. Now we have

$$\begin{aligned} 0 &= 0e_i \\ &= (x_1 + x_2 + \cdots + x_k)e_i \end{aligned}$$

$$\begin{aligned}
&= x_1 e_i + \cdots + x_i e_i + \cdots + x_k e_i \\
&= x_i
\end{aligned}$$

as we sought. □

Notice that the orthogonal idempotents associated with the decomposition of $M_n(F)$ that we gave in Example 4.2.3 are the matrix units $e_{11}, e_{22}, \dots, e_{nn}$.

Direct sum decompositions $M = M_1 \oplus M_2 \oplus \cdots \oplus M_k$ of a general R -module M can be shown to correspond to the orthogonal idempotent decompositions $1 = e_1 + e_2 + \cdots + e_k$ of the identity in the ring $\text{End}_R(M)$ of all R -endomorphisms of M for which $M_i = e_i(M)$. We will not be needing the correspondence in exactly this form, but rather in the form of two important sets of R -homomorphisms associated with any such direct sum decomposition: the *projections* $\pi_1, \pi_2, \dots, \pi_k$ (these are the orthogonal idempotent endomorphisms referred to) and the *injections* $\nu_1, \nu_2, \dots, \nu_k$.

Definition 4.2.5: Let

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_k$$

be a direct sum decomposition of an R -module M . For $i = 1, 2, \dots, k$, the *ith projection* is the R -endomorphism

$$\pi_i : M \rightarrow M_i, x \mapsto x_i$$

where x_i is the *ith* component of x in the decomposition. For $i = 1, 2, \dots, k$, the *ith injection* is the R -homomorphism

$$\nu_i : M_i \rightarrow M, x \mapsto x.$$

Notice that $\pi_i \nu_i$ is the identity mapping on M_i whereas $\pi_i \nu_j$ is the zero mapping on M_j when $i \neq j$. Also, each π_i is an idempotent endomorphism of M with $\pi_1 + \pi_2 + \cdots + \pi_k = 1_M$, the identity mapping on M .

A submodule A of the module M is called a **direct summand** of M if $M = A \oplus B$ for some submodule B of M . The complementary summand B is far from unique. For instance, in the case of a vector space, if we choose a basis \mathcal{B}_1 for A , and extend it to a basis \mathcal{B} for M , then the subspace B spanned by $\mathcal{B} \setminus \mathcal{B}_1$ is a complementary subspace of A . (Conversely, all complementary subspaces take this form.) On the other hand, any two complementary submodules of A must be isomorphic, because applying the Fundamental Theorem 4.1.9 to the projection $M = A \oplus B \rightarrow B$ yields $B \cong M/A$. Note also that if $M = M_1 \oplus M_2 \oplus \cdots \oplus M_k$, then each M_i is a direct summand of M .

Any module M has both itself and its zero submodule 0 as direct summands because $M = M \oplus 0$. If these are the only direct summands of a nonzero module M , then M is called an **indecomposable module**, since it then has only the trivial direct sum decomposition. For instance, the indecomposable vector spaces are the 1-dimensional ones, and the indecomposable finitely generated \mathbb{Z} -modules are the infinite cyclic groups and the finite groups of prime power order (by the fundamental theorem⁸ for finitely generated abelian groups). For a ring R regarded as a module over itself, being indecomposable is, by Proposition 4.2.4, equivalent to R having only the trivial idempotents 0 and 1 . In particular, every integral domain is indecomposable.

One of the most useful properties of direct sums is given in the next proposition. (It also holds for external direct sums, and in fact the property characterizes the external $M_1 \oplus M_2 \oplus \cdots \oplus M_k$ to within module isomorphism.)

Proposition 4.2.6

Suppose $M = M_1 \oplus M_2 \oplus \cdots \oplus M_k$ is a direct sum decomposition of an R -module M . Given any other R -module N and R -homomorphisms $f_i : M_i \rightarrow N$ for $i = 1, 2, \dots, k$, there is a unique R -homomorphism $f : M \rightarrow N$ that extends the f_i , that is, $f(x) = f_i(x)$ whenever $x \in M_i$. In terms of the injections $v_i : M_i \rightarrow M$, this says there is a unique f that makes the diagrams

$$\begin{array}{ccc} & & M \\ & \nearrow v_i & \vdots \\ M_i & \xrightarrow{f_i} & N \end{array} \quad \begin{array}{c} \\ \\ f \end{array}$$

commutative for each $i = 1, 2, \dots, k$.

Proof

From directness of the decomposition, each $x \in M$ can be written uniquely as $x = x_1 + x_2 + \cdots + x_k$ with each $x_i \in M_i$. We simply set

$$f(x) = f_1(x_1) + f_2(x_2) + \cdots + f_k(x_k).$$

Then f is a well-defined map and is easily checked to be an R -homomorphism that meets our requirements. (Of course, we had no choice in how to define f .) \square

The next two little lemmas will come in handy later. The first tells us that if A is a direct summand of a module M , then A is also a direct summand of any intermediate submodule C .

8. See, for example, Jacobson's *Basic Algebra I*, Chapter 3.

Lemma 4.2.7

Suppose A is a direct summand of a module M with $M = A \oplus B$. If C is a submodule of M containing A , then

$$C = A \oplus (B \cap C).$$

Proof

Let $D = B \cap C$. We need only check that $C = A + D$ and $A \cap D = 0$. From $M = A \oplus B$ we know $M = A + B$ and $A \cap B = 0$. So certainly $A \cap D = 0$. Let $c \in C$. Write $c = a + b$ for some $a \in A, b \in B$. Then, since $A \subseteq C$, we have $b = c - a \in D$. Thus, $c = a + b \in A + D$. This shows that $C = A + D$. \square

Lemma 4.2.8

Let $f : M \rightarrow N$ be a module homomorphism.

- (1) If $\ker(f)$ is a direct summand of M , then there is a homomorphism $k : \text{im}(f) \rightarrow M$ such that fk is the identity on $\text{im}(f)$.
- (2) Conversely, the existence of such a homomorphism k implies that $\ker(f)$ is a direct summand of M .

Proof

(1) Assume there is a submodule B of M with $M = \ker(f) \oplus B$. All part (1) of the lemma is saying is that the restriction g of f to B is an R -isomorphism from B onto $\text{im}(f)$, and then we can choose k to be the inverse mapping:

$$M = \left\{ \begin{array}{ccc} \ker(f) & \longrightarrow & 0 \\ \oplus & & \\ B & \xrightleftharpoons[k]{g=f|_B} & \text{im}(f) \end{array} \right.$$

We only need check that g is an onto mapping with zero kernel. Now any $m \in M$ can be written as $m = a + b$ with $a \in \ker(f)$ and $b \in B$, and so

$$f(m) = f(a) + f(b) = 0 + f(b) = g(b).$$

This shows g maps B onto $\text{im}(f)$. Also $\ker(g) = \ker(f) \cap B = 0$ from the directness of $M = \ker(f) \oplus B$. Thus, g is a module isomorphism.

(2) For the converse, suppose k exists, let $C = \text{im}(f)$, and let 1_C denote the identity mapping on C . Let $e = kf : M \rightarrow M$. Then e is an idempotent endomorphism of

M because $e^2 = kfkf = k1_Cf = kf = e$. By the same direct sum arguments used in the proof of Proposition 4.2.4, we have $M = e(M) \oplus (1 - e)M$. One easily checks $\ker(f) = (1 - e)M$. Thus, $\ker(f)$ is a direct summand of M . \square

4.3 FREE AND PROJECTIVE MODULES

In contrast to modules over a field F , modules over a general ring R can have quite complicated structures. The situation is somewhat improved when we restrict to certain classes of modules. Free and projective modules are among the better behaved, and results there more closely parallel their vector space counterparts. We briefly describe these modules in this section, principally in the case where they are also finitely generated.

The most useful fact concerning a vector space is the existence of a basis. We can formulate a basis for a general module as in the following definition. A **free R -module** is one that possesses a basis. To simplify our presentation, we'll only detail the case where the bases are finite. However, just as for vector spaces, there is a natural extension of the definition that accommodates infinite bases.

Definition 4.3.1: Let M be an R -module. A (finite) **basis** for M is a set $X = \{x_1, x_2, \dots, x_n\}$ of elements of M satisfying:

- (1) X generates M , that is, $M = Rx_1 + Rx_2 + \dots + Rx_n$.
- (2) X is linearly independent in the sense that if $r_1, r_2, \dots, r_n \in R$ with

$$r_1x_1 + r_2x_2 + \dots + r_nx_n = 0$$

then $r_1 = r_2 = \dots = r_n = 0$.

An equivalent statement to (1) and (2) (combined) is that each $m \in M$ can be written uniquely in the form

$$(*) \quad m = r_1x_1 + r_2x_2 + \dots + r_nx_n.$$

The map $f : R^n \rightarrow M$, $(r_1, r_2, \dots, r_n) \mapsto r_1x_1 + r_2x_2 + \dots + r_nx_n$ is then an R -module isomorphism from the external direct sum $R^n = R \oplus R \oplus \dots \oplus R$ of n copies of R (as a module over itself) onto M . Conversely, if there exists such an isomorphism, then the image of the standard basis $\{(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, 0, \dots, 1)\}$ is a basis for M .

Thus, from a module viewpoint, the finitely generated free modules are exactly those modules M isomorphic to a finite direct sum of copies of R .⁹

From a universal algebra point of view, a basis X for M can be characterized by the property that any function $f : X \rightarrow N$ from X into an arbitrary R -module N can be uniquely extended to a module homomorphism $\bar{f} : M \rightarrow N$. That is, in terms of the inclusion mapping $i : X \rightarrow M$, there is a unique R -homomorphism \bar{f} that makes the diagram

$$\begin{array}{ccc} & & M \\ & \nearrow i & \vdots \bar{f} \\ X & \xrightarrow{f} & N \end{array}$$

commutative. Of course, in view of $(*)$ we have no choice but to take

$$\bar{f}(m) = r_1 f(x_1) + r_2 f(x_2) + \cdots + r_n f(x_n).$$

Remark 4.3.2

Suppose R is a commutative ring and M is a free R -module with a basis $X = \{x_1, x_2, \dots, x_n\}$. Given an R -module endomorphism $f : M \rightarrow M$, we can form the $n \times n$ **matrix** $A = (a_{ij})$ **of f relative to X** in exactly the same way as for linear transformations. Namely, the a_{ij} are the members of R uniquely determined from the relations

$$f(x_j) = \sum_{i=1}^n a_{ij} x_i.$$

The correspondence $f \mapsto A$ is then a ring isomorphism of $\text{End}_R(M)$ onto $M_n(R)$. \square

All vector spaces over a field F are free as F -modules. All finitely generated torsion-free abelian groups are free as \mathbb{Z} -modules. However, for general rings, free modules are pretty thin on the ground. Even a direct summand of a free module need not be free. For instance, let R be the ring $M_n(F)$ of $n \times n$ matrices over a field F , and let $M = R$ as a module over itself. Let $P = Re_{11}$ be the left ideal of R consisting of all matrices with zero columns except possibly for the first. Then P is a direct summand of M , and P is not free when $n > 1$, otherwise the vector space dimension of P would be a multiple of $\dim R = n^2$.

9. However, the number of copies of R involved need not be an invariant for M , unless R is a commutative ring or some other “nice” ring. In other words, two different bases need not have the same number of elements.

But $\dim P = n$. However, all is not lost in this type of example. Any module P that is (isomorphic to) a direct summand of a free module has a rather nice property, equivalent to being *projective* in the following sense.

Definition 4.3.3: An R -module P is called **projective** if, given an epimorphism $g : M \rightarrow N$ of an R -module M onto an R -module N , then every homomorphism $f : P \rightarrow N$ can be lifted to a homomorphism $h : P \rightarrow M$ in the sense that $f = gh$ ($= g \circ h$). In other words, the diagram

$$\begin{array}{ccc} & P & \\ h \swarrow & \downarrow f & \\ M & \xrightarrow{\quad g \quad} & N \end{array}$$

commutes.

Proposition 4.3.4

A finitely generated R -module P is projective if and only if it is isomorphic to a direct summand of a finitely generated free R -module.

Proof

We prove the “if part” in two stages: (1) A free module is projective, and (2) a direct summand of a projective module is projective.

(1) Let Q be a free module with a basis X . Given an epimorphism $g : M \rightarrow N$ and a homomorphism $f : Q \rightarrow N$, we need to produce a homomorphism $h : Q \rightarrow M$ to make the diagram

$$\begin{array}{ccc} & Q & \\ h \swarrow & \downarrow f & \\ M & \xrightarrow{\quad g \quad} & N \end{array}$$

commutative. Since g is onto, for each $x \in X$ we can choose an element m of M with $f(x) = g(m)$. Since X is a basis for Q , the function $X \rightarrow M$, $x \mapsto m$ can be extended to an R -homomorphism $h : Q \rightarrow M$. Now the homomorphisms gh and f certainly agree on X , whence they must agree on Q because X generates Q . This establishes that Q is projective.

(2) Next, suppose P is a direct summand of a projective module Q , say

$$Q = P \oplus B.$$

Let g and f be given as in Definition 4.3.3. We have to produce a homomorphism h that makes the diagram

$$\begin{array}{ccc} & P & \\ h \swarrow & \downarrow f & \\ M & \xrightarrow{g} & N \end{array}$$

commutative. Let $\pi : Q = P \oplus B \rightarrow P$ and $\nu : P \rightarrow Q$ be, respectively, the projection and injection homomorphisms associated with the first summand P (see Definition 4.2.5). Since Q is projective, there is a homomorphism $k : Q \rightarrow M$ making the diagram

$$\begin{array}{ccc} & Q & \\ & \downarrow \pi & \\ k \swarrow & P & \\ & \downarrow f & \\ M & \xrightarrow{g} & N \end{array}$$

commutative. That is, $gk = f\pi$. Now let $h = k\nu : P \rightarrow M$. Since $\pi\nu = 1_P$, the identity mapping on P , we have

$$gh = gk\nu = f\pi\nu = f.$$

This shows P is projective.

The “only if” part of the proposition follows as a corollary to a more general result on projective modules, which we shall record separately as our next theorem. To see this, note that every finitely generated module M is a homomorphic image of some finitely generated free module Q : for if M is generated by $S = \{m_1, m_2, \dots, m_n\}$, choose a free module Q with a basis $X = \{x_1, x_2, \dots, x_n\}$ and use freeness to extend the function $X \rightarrow S$, $x_i \mapsto m_i$ to a homomorphism $f : Q \rightarrow M$.¹⁰ The map must be onto because the image of f includes the generators S for M . \square

Theorem 4.3.5

A module P is projective if and only if every epimorphism $f : M \rightarrow P$ from a general module M onto P **splits**, in the sense that $\ker(f)$ is a direct summand of

10. The same argument shows that every R -module is a homomorphic image of some free module, not necessarily possessing a finite basis.

M , equivalently, there is a homomorphism $k : P \rightarrow M$ such that $fk = 1_P$, the identity mapping on P . (Then k maps P isomorphically onto the direct summand $(kf)(M)$ of M .)

$$M = \left\{ \begin{array}{ccc} \ker(f) & \longrightarrow & 0 \\ & \oplus & \\ \operatorname{im}(f) & \xrightleftharpoons[k]{f} & P \end{array} \right.$$

Proof

If P is projective, the existence of k comes from applying that property to the diagram

$$\begin{array}{ccc} & P & \\ k \swarrow & \downarrow 1_P & \\ M & \longrightarrow & P \end{array}$$

An appeal to Lemma 4.2.8 gives the equivalence with $\ker(f)$ being a direct summand. Since $fk = 1_P$, we see that kf is an idempotent endomorphism of M , whence

$$M = (kf)(M) \oplus (1 - kf)M = k(P) \oplus \ker(f).$$

Conversely, if P has the stated property, we can use that on an epimorphism $f : M \rightarrow P$ from some free module M onto P . Then P is isomorphic to a direct summand of M , whence P is projective by Proposition 4.3.4. \square

The property in the above theorem, that if P is a homomorphic image of some module M , then P must be (isomorphic to) a direct summand of M , is one of the most cited properties of a projective module P .¹¹

The class of projective modules is quite large. For example, all modules over the ring $M_n(F)$ turn out to be projective. But we are not quite done yet with our generalizations. In providing a module setting for the Weyr form, it is just as easy to work with *quasi-projective* modules. Hang on, we hear the reader say, isn't that what is wrong with modern ring theory—generalizations for the sake of generalization, coming up with an obscure property (P) of

11. It also prompts the maxim *The best things in life are free; the next best are projective.*

which there appear to be no interesting examples, and then giving a theorem showing (P) is equivalent to 15 even more obscure conditions?¹² Ah, but quasi-projective modules also include a very nice class of natural modules, the so-called *semisimple* (or *completely reducible*) modules, which are not always projective. A **quasi-projective** module P is one satisfying the conditions in Definition 4.3.3 in the special case $M = P$: given an epimorphism $g : P \rightarrow N$ and a homomorphism $f : P \rightarrow N$, there exists an endomorphism $h : P \rightarrow P$ such that the diagram

$$\begin{array}{ccc} & P & \\ h \swarrow & \downarrow f & \\ P & \xrightarrow{g} & N \end{array}$$

commutes. A **semisimple** (or **completely reducible**) module M is one that is a sum of simple submodules (and in turn this is equivalent to M being a direct sum of some family of simple modules). Here we need not require the family of simple submodules involved to be finite. In general, given a family $\{M_i : i \in I\}$ of submodules, their **sum** $\sum_{i \in I} M_i$ is the submodule consisting of all elements of the form $\sum_{i \in I} m_i$ where $m_i \in M_i$ and $m_i = 0$ for almost all $i \in I$. The characteristic property of a semisimple module M is that every submodule is a direct summand. The interested reader can consult Jacobson's *Basic Algebra II*, Theorem 3.10, for details. Modulo this, we can establish the following:

Proposition 4.3.6

Every semisimple module M is quasi-projective.

Proof

Assume M is semisimple. Let $f, g : M \rightarrow N$ be R -homomorphisms with g an onto map. We require an endomorphism $h : M \rightarrow M$ to complete the commutative picture

$$\begin{array}{ccc} & M & \\ h \swarrow & \downarrow f & \\ M & \xrightarrow{g} & N \end{array}.$$

12. The first author believes that the literature does contain too many of these “(a), (b), (c), . . . , (z) theorems.” Good ring theory isn’t about this.

By semisimplicity of M , we know $\ker(g)$ is a direct summand of M . Hence, by Lemma 4.2.8 (1), and the fact that $\operatorname{im}(g) = N$, there is a homomorphism $k : N \rightarrow M$ such that $gk = 1_N$, the identity mapping on N . Let $h = kf : M \rightarrow M$. We have $gh = gkf = f$, which establishes that M is quasi-projective. \square

To fulfill our campaign promise, we need to exhibit a semisimple module that is not projective. That is easy. Even simple modules need not be projective. For instance, a cyclic group G of prime order is a simple \mathbb{Z} -module but not projective, otherwise by Proposition 4.3.4 it would be isomorphic to a direct summand of a free \mathbb{Z} -module. But a free \mathbb{Z} -module is in particular a torsion-free group, so G can't sit inside it even as a subgroup.

The key properties of quasi-projective modules that we require later in our Weyr form derivation are contained in the following theorem and proposition, whose projective versions we have seen in the proof of Proposition 4.3.4 and (the statement of) Theorem 4.3.5:

Theorem 4.3.7

Let Q be a quasi-projective module (over an arbitrary ring). Then:

- (1) *Any direct summand A of Q is quasi-projective.*
- (2) *If $f : Q \rightarrow Q$ is an endomorphism such that $\operatorname{im}(f)$ is a direct summand of Q , then $\ker(f)$ is a direct summand of Q .¹³*

Proof

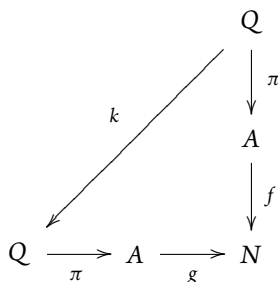
(1) The usual setup applies. We need to supply the dotted homomorphism for commutativity of the following diagram, given homomorphisms f and g with g onto.

$$\begin{array}{ccc} & & A \\ & \nearrow h & \downarrow f \\ A & \xrightarrow{g} & N \end{array}$$

Assume $Q = A \oplus B$ and let $\pi : Q = A \oplus B \rightarrow A$ and $\nu : A \rightarrow Q$ be, respectively, the projection and injection homomorphisms associated with the first summand A . Since Q is quasi-projective, there is a homomorphism $k : Q \rightarrow Q$ making the

13. However, this property (2) does not characterize quasi-projectivity, as demonstrated by Xue in 1993.

diagram

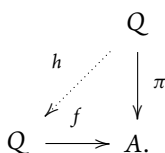


commutative. That is, $g\pi k = f\pi$. Let $h = \pi k v : A \rightarrow A$. Then, noting that πv is the identity mapping on A , we have

$$gh = g\pi k v = f\pi v = f,$$

which shows A is quasi-projective.

(2) Let $A = \text{im}(f)$ and assume $Q = A \oplus B$. Using the same associated maps π and v as in (1), since Q is quasi-projective, there is a homomorphism $h : Q \rightarrow Q$ giving commutativity of



That is, $fh = \pi$. Let $k = hv : A \rightarrow Q$. Then

$$fk = fhv = \pi v = 1_A.$$

By Lemma 4.2.8, this implies that $\ker(f)$ is a direct summand of Q . □

Since there are quasi-projective modules that are not projective, we can't expect the splitting property in Theorem 4.3.5 to hold for all quasi-projectives (try it on the canonical epimorphism $\mathbb{Z} \rightarrow \mathbb{Z}/p\mathbb{Z}$ for a prime p). However, the following weaker property is all that we will require later.

Proposition 4.3.8

Let $Q = A \oplus B$ be a direct sum decomposition of a quasi-projective module Q . Then every epimorphism $g : B \rightarrow A$ splits.

Proof

We can extend g to an endomorphism $f : Q \rightarrow Q$ by letting f map A to 0.

$$Q = \left\{ \begin{array}{ccc} A & \longrightarrow & 0 \\ \oplus & & \\ B & \xrightarrow{g} & A \end{array} \right.$$

Now $\text{im}(f) = A$ is a direct summand of Q , so $\ker(f)$ is a direct summand of Q by Theorem 4.3.7 (2). But $\ker(f) = A \oplus \ker(g)$ and therefore $\ker(g)$ is a direct summand of Q . Finally, by Lemma 4.2.7, we know $\ker(g)$ must be a direct summand of B . Thus, g splits. \square

4.4 VON NEUMANN REGULARITY

The concept of a *von Neumann regular* element of a ring is a very useful one. It was introduced in the mid 1930s by John von Neumann, one of the great mathematicians¹⁴ of the 20th century, in connection with his work on algebras of operators on Hilbert spaces. A special case of the notion later became popular for $n \times n$ matrices over the reals or complexes through the so-called Moore–Penrose inverse A^+ of a matrix A . That notion continues to be important in linear algebra and applications, through its connection with the optimal least squares solution $\bar{x} = A^+b$ of an inconsistent linear system $Ax = b$. However, the name that is most deservedly associated with the promotion of “*things von Neumann regular*” is Kenneth R. Goodearl. Through his fine book *Von Neumann Regular Rings*, Goodearl presented the theory in a very coherent way in the late 1970s (2nd edition, 1991). The book inspired others to take up the study, which continues to this day. (However, of the 57 Open Problems posed by Goodearl in the first edition, 21 of them still remain completely open, and 31 not fully resolved, as of 2010.) In this section, we develop the basic properties of von Neumann regular elements in a completely general ring R .

Definition 4.4.1: Let R be a ring. An element $a \in R$ is said to be (von Neumann) **regular** if there exists an element $b \in R$ such that

$$a = aba.$$

14. See the biographical note at the end of this chapter.

Any such b is called a **quasi-inverse** of a . The ring R is called (von Neumann) **regular** if all its elements are regular.¹⁵

Here is the most important class of examples for our later Weyr form connection.

Example 4.4.2

The ring $M_n(F)$ of $n \times n$ matrices over a field F is regular. To show this, rather than work directly with the matrices, it is easier and more instructive to work with the isomorphic ring R of linear transformations of an n -dimensional vector space V over F . (If one prefers, take $V = F^n$ to be the space of $n \times 1$ column vectors and identify an $n \times n$ matrix with its left multiplication map on V .) Fix $a \in R$. Let $W = \ker(a)$ and $Z = \text{im}(a)$. Choose complementary subspaces X of W , and Y of Z . Thus, $V = W \oplus X = Y \oplus Z$. In terms of the diagram

$$V = \left\{ \begin{array}{ccc} W & \xrightarrow{0} & Y \\ \oplus & & \oplus \\ X & \begin{array}{c} \xrightarrow{a} \\ \xleftarrow{b} \end{array} & Z \end{array} \right.$$

the forward arrows describe the action of a and the backward arrow depicts the action on Z of our proposed quasi-inverse b of a . Since a is mapping X isomorphically onto Z , we can let b act on Z as the inverse of this mapping to give $a = aba$ on X . Now extending b to a linear transformation of V in any old

15. The potential for a conflict of terminology exists here with the well-established term of a “von Neumann algebra,” which is a weakly closed self-adjoint algebra of bounded operators on a Hilbert space. But as rings, von Neumann algebras are not (outside the finite-dimensional case) von Neumann regular rings. However, every “finite” von Neumann algebra can be canonically embedded in a regular ring whose principal left ideals co-ordinatize the lattice of projections of the von Neumann algebra. There is also a more recent connection of von Neumann algebras, more generally of the C^* -algebras of “real rank zero,” with the so-called *exchange rings*, which were introduced by R. Warfield in the 1970s, and generalize regular rings. (C^* -algebras are Banach algebras with an involution $*$ satisfying $\|xx^*\| = \|x\|^2$.) The finitely generated projective modules over an exchange ring have a lot in common with those of a regular ring. Every C^* -algebra A of real rank zero is an exchange ring. This connection was established by Ara, Goodearl, O’Meara, and Pardo in 1998. It allows certain questions about the lattice of projections of A to be re-evaluated in terms of finitely generated projective modules over an exchange ring.

There is also a conflict of terminology with “regular rings” in the setting of commutative noetherian rings. We won’t expand on this because it is unrelated to what we are doing. “Regular” must be one of the most overused adjectives in mathematics.

way (i.e., an arbitrary linear action on Y) will produce a quasi-inverse of a . This is because for $v \in V$, we have $v = w + x$ for some $w \in W$, $x \in X$ and

$$aba(v) = aba(w + x) = aba(w) + aba(x) = aba(x) = a(x) = a(v)$$

showing $a = aba$.

In actual fact, all the quasi-inverses c of a arise in this manner, through other choices of complements X and Y . For if $a = aca$, let $f = ca$, an idempotent transformation of V for which $\ker(a) = (1 - f)(V)$. Now taking $X_1 = f(V)$ and $Y_1 = Y$, we have that X_1 and Y_1 are complements of $W = \ker(a)$ and $Z = \text{im}(a)$, respectively, and the action of c on Z is the inverse of the restriction of a mapping X_1 isomorphically onto Z : for if we let $x = f(u) \in X_1$ (for some $u \in V$) and $z = a(v) \in Z$ (for some $v \in V$), we have

$$ca(x) = caf(u) = caca(u) = ca(u) = f(u) = x,$$

$$ac(z) = aca(v) = a(v) = z.$$

The reason why we haven't bothered choosing a new Y is that the action of c on Y is arbitrary anyway. \square

Two sensible choices for extending b from Z to V in Example 4.4.2 present themselves:

- (1) Let b map Y isomorphically onto W . This is possible because

$$\dim Y = \dim V - \dim Z = \dim V - \text{rank}(a) = \text{nullity}(a) = \dim W.$$

Then b is invertible and is accordingly referred to as a **unit-quasi-inverse** of a .

- (2) Let b map Y to zero. In this case, we have

$$a = aba \text{ and } b = bab.$$

In other words, a is also a quasi-inverse of b . In this case, b is referred to as a **generalized inverse** (or **pseudo-inverse**) of a . Notice that because the rank of a product is at most the rank of each of its factors, an equation $a = aba$ implies the rank of a quasi-inverse b is at least the rank of a . Another way of thinking of generalized inverses of a is that they are precisely the quasi-inverses having the same rank as a (the sufficiency of this condition is not hard to prove), the opposite extreme to a unit-quasi-inverse, which has the maximum rank possible.

Of course, if a happens to be invertible, its only quasi-inverse is its inverse. But when a is not invertible, there are many choices (infinitely many if F is infinite) for quasi-inverses, both in the choices for the complementary subspaces X and Y and for the action of the quasi-inverse on Y . Note however, that in our construction of a generalized inverse b , it is completely determined once X and Y are chosen, since then $X = \text{im}(b)$ and we have chosen $Y = \ker(b)$. This observation is worth a formal note.

Proposition 4.4.3

Let $a : V \rightarrow V$ be a fixed linear transformation with kernel W and image Z .

- (1) *There is a one-to-one correspondence between the generalized inverses of a and pairs (X, Y) of complementary subspaces X of W , and Y of Z : given a generalized inverse b , take $X = \text{im}(b)$ and $Y = \ker(b)$. The inverse correspondence is that, given X and Y , let b act as the inverse of a mapping X onto Z , and let b map Y to 0.*
- (2) *If (X, Y) and (X_1, Y_1) are two such pairs of complementary subspaces with associated generalized inverses b and c for a , then*

$$c = \pi_1 b \pi_2,$$

where π_1 and π_2 are the projections

$$\pi_1 : W \oplus X_1 \rightarrow X_1, \quad \pi_2 : Y_1 \oplus Z \rightarrow Z.$$

Proof

(1) If b is a generalized inverse of a , let $e = ab$ and $f = ba$. Then e and f are idempotents with

$$\ker(a) = (1 - f)(V), \quad \text{im}(a) = e(V)$$

$$\ker(b) = (1 - e)(V), \quad \text{im}(b) = f(V).$$

Let $X = f(V)$ and $Y = (1 - e)(V)$. Inasmuch as $V = g(V) \oplus (1 - g)(V)$ for any idempotent transformation g , we have $V = W \oplus X = Y \oplus Z$. Moreover, by the same argument used in Example 4.4.2, from $a = aba$ and $b = bab$ we see that a maps X isomorphically onto $Z = \text{im}(a)$ while b undoes this action. It follows that the indicated correspondence is one to one.

(2) Let π be the projection

$$\pi : W \oplus X \rightarrow X.$$

Note that π and π_1 are inverse mappings when restricted to X_1 and X , respectively. Therefore we have a commutative diagram

$$\begin{array}{ccccc}
 & & & & V \\
 & & & & \downarrow \pi_2 \\
 & & c & & \\
 & \swarrow & & \searrow & \\
 X_1 & \xrightleftharpoons[\pi_1]{} & X & \xrightleftharpoons[b]{} & Z
 \end{array}$$

in which the composition of the bottom forward maps is the action of a on X_1 , while the composition of the backward ones is its inverse. Thus, (2) holds. \square

Next we give an important ring-theoretic characterization of when an element a of a ring R is regular: the principal left ideal $Ra = \{ra : r \in R\}$ must be generated by an idempotent.

Proposition 4.4.4

Let R be a ring and let $a \in R$. Then the following are equivalent:

- (1) a is a regular element of R .
- (2) $Ra = Re$ for some idempotent e of R .
- (3) Ra is a direct summand of R as a left R -module.

Proof

(1) \iff (2). If a is regular, then $a = aba$ for some $b \in R$. Let $e = ba$. We have

$$e^2 = (ba)(ba) = b(aba) = ba = e$$

whence e is idempotent. Clearly, $Re \subseteq Ra$ because $e \in Ra$. But $a = aba = ae \in Re$ so the reverse containment also holds. Thus, $Ra = Re$. Conversely, if $Ra = Re$ for some idempotent e , then $e = ba$ and $a = re$ for some $b, r \in R$. Now

$$a = re = re^2 = ree = ae = aba$$

and so a is regular with b as a quasi-inverse.

(2) \iff (3). We know this from Proposition 4.2.4. \square

Of course, the principal right ideal versions of the proposition also hold. It turns out that if R itself is a regular ring, then all finitely generated left (right) ideals are generated by an idempotent, whence projective as left (right)

R -modules. Moreover, every finitely generated projective left R -module is (isomorphic to) a finite direct sum of principal left ideals. We won't be making use of this property but mention it in passing because it is one of the more useful facts about regular rings.

Remarks 4.4.5

- (1) If we know one idempotent generator e of the principal left ideal Ra , we know them all. The general one is $h = e + (1 - e)xe$, where $x \in R$ is arbitrary.
- (2) If a is a regular element of R and b is a quasi-inverse of a , then the element $g = bab$ is a generalized inverse of a . For we have

$$\begin{aligned} a &= aba = (aba)ba = a(bab)a = aga, \\ gag &= (bab)a(bab) = babab = bab = g. \end{aligned}$$

- (3) Suppose a is a regular element of R . Then there is a 1-1 correspondence between generalized inverses b of a and pairs of idempotents (e, f) of R which generate the same principal left and right ideals as a :

$$Ra = Re \text{ and } aR = fR.$$

In one direction, given b let $e = ba$ and $f = ab$. In the other direction, given e and f , let b be the unique element of eRf satisfying $e = ba$. We leave the details as an exercise. \square

When we come to relate the Weyr form to von Neumann regularity, the relevant ring is the ring $\text{End}_R(M)$ of all R -endomorphisms of an appropriate left R -module M , but over a completely general ring R . We need, therefore, to recognize when an endomorphism $f : M \rightarrow M$ is regular. Here is how:¹⁶

Proposition 4.4.6

An endomorphism $a : M \rightarrow M$ of an R -module M is regular in $\text{End}_R(M)$ exactly when $\ker(a)$ and $\text{im}(a)$ are direct summands of M . If M is quasi-projective, then regularity of a is equivalent to just $\text{im}(a)$ being a direct summand of M .

Proof

First, assume a is regular, say $a = aba$ for some $b \in \text{End}_R(M)$. Let $e = ba$ and $f = ab$. Then e and f are idempotent endomorphisms and one easily checks that

$$\ker(a) = (1 - e)M, \text{ im}(a) = f(M).$$

16. This useful result appears in the 1971 paper by R. Ware, but was probably known much earlier than that.

For instance, if $x \in \ker(a)$, then $x = e(x) + (1 - e)(x) = b(a(x)) + (1 - e)(x) = (1 - e)(x) \in (1 - e)M$. Since e and f are idempotent, we have direct sum decompositions $M = e(M) \oplus (1 - e)M = f(M) \oplus (1 - f)M$, whence $\ker(a)$ and $\text{im}(a)$ are direct summands of M .

The converse is just an extension of the argument used in Example 4.4.2, where $M = V$ is regarded as a module over the field F . Assume $W = \ker(a)$ and $Z = \text{im}(a)$ are direct summands of M , say $M = W \oplus X = Y \oplus Z$ for some submodules X and Y of M . We can produce a quasi-inverse b of a by exactly the earlier argument. Namely, since a maps X isomorphically onto Z (we are using the direct sum decomposition $M = W \oplus X$ for this assertion—see Lemma 4.2.8 and its proof), we can let b act on Z as the inverse of this map, and on Y let b be any homomorphism into M (for instance, the zero mapping).

$$M = \left\{ \begin{array}{ccc} W & \xrightarrow{0} & Y \\ \oplus & & \oplus \\ X & \xrightleftharpoons[b]{a} & Z \end{array} \right.$$

Our use of the direct sum decomposition $M = Y \oplus Z$ is that putting the two parts of b together makes sense: any $m \in M$ can be written uniquely as $m = y + z$ for some $y \in Y$, $z \in Z$ and we let

$$b(m) = b(y) + b(z).$$

To check that $a = aba$, it is enough to check that the two maps a and aba agree on W and X (since $M = W + X$). But both are zero on W , while on X we just use the fact that b is the inverse map of a . Thus, a is regular.

The final assertion of the proposition follows from Theorem 4.3.7 (2). \square

Remarks 4.4.7

- (1) For the left R -module $M = R$, the ring $\text{End}_R(M)$ of R -endomorphisms is anti-isomorphic to R itself: given $a \in R$, let $\theta_a: M \rightarrow M$ be the right multiplication map $m \mapsto ma$ of M by a . Then $\theta: R \rightarrow \text{End}_R(M)$, $a \mapsto \theta_a$ is a bijection (the inverse map is $f \mapsto f(1)$), which preserves addition but switches products: $\theta(ab) = \theta(b)\theta(a)$. This is because of our choice of function composition. Clearly, a is regular if and only if θ_a is regular. Since M is projective, the test for the latter is whether $\text{im}(\theta_a)$ is a direct summand of M . But $\text{im}(\theta_a) = Ra$. Thus, we can view the equivalence of (1) and (3) in Proposition 4.4.4 as a special case of Proposition 4.4.6.

- (2) In general, having $\ker(a)$ as a direct summand of M is not enough for regularity of an endomorphism. Witness (1) in the case $M = R$ where R is an integral domain, and a is a nonzero nonunit. Then $\ker(\theta_a) = 0$, which is a direct summand, but θ_a is not regular (because a is not, since otherwise it would be zero or invertible).
- (3) Further to (1), we note that if Re and Rf are principal left ideals of a ring R generated by idempotents e and f , then there is an additive group isomorphism from $eRf = \{erf : r \in R\}$ onto the group $\text{Hom}(Re, Rf)$ of all R -homomorphisms from Re to Rf : given $a \in eRf$, let $\theta_a: Re \rightarrow Rf$ be right multiplication by a . Another way of viewing the generalized inverse b associated with idempotent generators e and f of Ra and aR in Remark 4.4.5 (3) is that θ_b is the inverse of the module homomorphism θ_a . \square

Corollary 4.4.8

The ring $\text{End}_R(M)$ of all R -endomorphisms of a semisimple module M is regular. In particular, the ring of all linear transformations of a vector space (not necessarily finite-dimensional) is regular.

Proof

All submodules of a semisimple module are direct summands. In particular, the kernel and image of an endomorphism are direct summands. \square

4.5 COMPUTING QUASI-INVERSES

Our discussions so far really only demonstrate the existence of quasi-inverses, not how to compute them. Even in the special case of an invertible $n \times n$ matrix, we have said nothing. No mention of computing its inverse by row operations, for instance, or the formula $A^{-1} = (1/\det A) \text{adj } A$. One gets the feeling that maybe one reason why there are many unsolved problems in regular rings is that not enough attention has been paid to the way quasi-inverses can be constructed. We really don't fully understand, at a computational level, even quasi-inverses of $n \times n$ matrices over a field.¹⁷ To practice what we preach, let's actually compute some quasi-inverses of simple matrices. Those readers

17. Currently, one of the most important open problems in regular rings is the so-called Separativity Problem, formulated by Ara, Goodearl, O'Meara, and Pardo in 1998. It asks whether all regular rings are "separative," which is a certain cancellation property of their finitely generated projective modules with respect to direct sums and isomorphism. A number of outstanding open problems in regular rings have positive answers for separative regular rings, but the evidence strongly points to the Separativity Problem having a negative answer. However, no nonseparative regular rings have yet been constructed (partly because separativity is preserved in standard constructions, such as extensions of ideals by factor rings). It is possible to formulate the Separativity Problem entirely in terms of a certain "uniform diagonalization formula" involving 2×2 matrices over the ring $M_n(F)$ for a field F . The formula must be *independent* of n but

who do not feel the need for such computations can safely proceed to the next section.

The following two lemmas will aid our calculations. The first shows how to compute all the complements of a given subspace if we already know one.

Lemma 4.5.1

Let W be a fixed m -dimensional subspace of an n -dimensional vector space V . Suppose X is one complementary subspace of W (that is, $V = W \oplus X$). Choose a basis $\mathcal{B} = \{x_1, x_2, \dots, x_n\}$ for V in which the first m vectors span W and the last $n - m$ span X . Then the general complement of W is precisely a subspace C of V that is spanned by vectors whose coordinate vectors relative to \mathcal{B} are the columns of some $n \times (n - m)$ matrix of the form

$$\begin{bmatrix} P \\ I \end{bmatrix},$$

where P is $m \times (n - m)$ and I is the $(n - m) \times (n - m)$ identity matrix. Different choices of P yield different complements C .

Proof

Given a complement C , choose an isomorphism $s : X \rightarrow C$ and let $t : V \rightarrow V$ be the isomorphism

$$V = \begin{cases} W & \xrightarrow{1} & W \\ \oplus & & \oplus \\ X & \xrightarrow{s} & C \end{cases}$$

Relative to \mathcal{B} , the matrix of t takes the form

$$[t]_{\mathcal{B}} = \left[\begin{array}{c|c} I_m & P \\ \hline 0 & Q \end{array} \right],$$

where Q is an invertible $(n - m) \times (n - m)$ matrix. We can arrange for Q to be I_{n-m} by replacing t with its composition (firstly) with $u : V \rightarrow V$ such that

$$[u]_{\mathcal{B}} = \left[\begin{array}{c|c} I_m & 0 \\ \hline 0 & Q^{-1} \end{array} \right].$$

hold for all quasi-inverses of matrices in $M_n(F)$. This highlights the need to understand more fully quasi-inverses in this matrix setting. Essentially the module-theoretic derivation of the Weyr form that we present in Section 4.8 evolved from a study of “uniform diagonalization” by Beidar, O’Meara, and Raphael in 2004.

Since $t(X) = C$, we have that C is spanned by the vectors whose coordinate vectors relative to \mathcal{B} are the columns of $\begin{bmatrix} P \\ I \end{bmatrix}$.

This process is reversible, simply by defining t via the matrix

$$[t]_{\mathcal{B}} = \left[\begin{array}{c|c} I_m & P \\ \hline 0 & I_{n-m} \end{array} \right].$$

The image C of X under t must be a complement of W because t is an isomorphism fixing W , whence will map complements of W to other complements of W . Finally, to see that different choices of P will give different complements, just observe that the matrices

$$\begin{bmatrix} P \\ I \end{bmatrix}$$

will have different column spaces (by the same argument showing two different matrices in reduced row-echelon form have different row spaces). \square

Lemma 4.5.2

In the notation of Lemma 4.5.1, suppose C is the complement of W associated with the matrix

$$\begin{bmatrix} P \\ I \end{bmatrix}.$$

Let π_C and π_W be the projections

$$\pi_C : W \oplus C \rightarrow C, \quad w + c \mapsto c$$

$$\pi_W : W \oplus C \rightarrow W, \quad w + c \mapsto w.$$

Then their matrices relative to \mathcal{B} are

$$[\pi_C]_{\mathcal{B}} = \left[\begin{array}{c|c} 0 & P \\ \hline 0 & I_{n-m} \end{array} \right], \quad [\pi_W]_{\mathcal{B}} = \left[\begin{array}{c|c} I_m & -P \\ \hline 0 & 0 \end{array} \right].$$

Proof

Let π_X be the projection $W \oplus X \rightarrow X$ onto X . Let $t : V \rightarrow V$ be the isomorphism determined by

$$[t]_{\mathcal{B}} = \left[\begin{array}{c|c} I_m & P \\ \hline 0 & I_{n-m} \end{array} \right].$$

Then $\pi_C = t\pi_X t^{-1}$, whence

$$\begin{aligned} [\pi_C]_{\mathcal{B}} &= [t]_{\mathcal{B}} [\pi_X]_{\mathcal{B}} [t]_{\mathcal{B}}^{-1} \\ &= \left[\begin{array}{c|c} I & P \\ \hline 0 & I \end{array} \right] \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & I \end{array} \right] \left[\begin{array}{c|c} I & -P \\ \hline 0 & I \end{array} \right] \\ &= \left[\begin{array}{c|c} 0 & P \\ \hline 0 & I \end{array} \right]. \end{aligned}$$

The form of $[\pi_W]_{\mathcal{B}}$ follows from the fact that $\pi_W = 1_V - \pi_C$. □

Example 4.5.3

Let F be a field of characteristic zero. What are the generalized inverses C of the matrix

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 2 & 3 \\ -1 & -2 & 2 \end{bmatrix}?$$

We could find all such C by solving the 18 equations in 9 variables that arise from the matrix equations $A = ACA$ and $C = CAC$. Not a good idea.¹⁸ Instead, let's rework specifically the steps in Example 4.4.2 and Proposition 4.4.3. Thus, we can take $V = F^3$, associate with A the linear transformation $a : V \rightarrow V$ given by left multiplication by A , and aim to construct all the generalized inverses c of a from one particular generalized inverse b . We can then translate back to matrices C once we know the action of c on the standard basis $\mathcal{B} = \{v_1, v_2, v_3\}$ for V , by taking C to be its matrix relative to this basis.

By elementary row operations, we have

$$A \longrightarrow \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

whence

$$x_1 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$$

18. For a given $A \in M_n(F)$, it is not hard to solve $ACA = A$ for C in terms of one known generalized inverse B of A ; one can show that $C = B + (I - BA)D + E(I - AB)$ where $D, E \in M_n(F)$ are arbitrary. However, it is not easy to recognize from this description which of these quasi-inverses are in fact generalized inverses of A .

forms a basis for $W = \ker(a)$. Letting $x_2 = v_2$, $x_3 = v_3$, we have a basis

$$\mathcal{B}_1 = \{x_1, x_2, x_3\}$$

for V , so we can complement W with the subspace $X = \langle x_2, x_3 \rangle$. Let

$$z_2 = a(x_2) = \begin{bmatrix} 2 \\ 2 \\ -2 \end{bmatrix}, \quad z_3 = a(x_3) = \begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix}.$$

Then z_2, z_3 form a basis for $Z = \text{im}(a)$. Taking $z_1 = v_1$ then gives us our third basis for V ,

$$\mathcal{B}_2 = \{z_1, z_2, z_3\},$$

for which the subspace $Y = \langle z_1 \rangle$ complements the subspace Z .

As in Example 4.4.2, we can construct one generalized inverse b of a by letting b be the transformation determined by

$$b(z_1) = 0, \quad b(z_2) = x_2, \quad b(z_3) = x_3.$$

Since

$$\begin{aligned} v_1 &= z_1 \\ v_2 &= -\frac{2}{5}z_1 + \frac{1}{5}z_2 + \frac{1}{5}z_3 \\ v_3 &= \frac{3}{5}z_1 - \frac{3}{10}z_2 + \frac{1}{5}z_3, \end{aligned}$$

we have

$$[b]_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{5} & -\frac{3}{10} \\ 0 & \frac{1}{5} & \frac{1}{5} \end{bmatrix}.$$

Now to the construction of the other generalized inverses of a , using Proposition 4.4.3. Relative to \mathcal{B}_1 a general complement X_1 of W is determined by a 3×2 matrix

$$\begin{bmatrix} P \\ I_2 \end{bmatrix},$$

where

$$P = \begin{bmatrix} \alpha & \beta \end{bmatrix}$$

is arbitrary, and the projection $\pi_1 : W \oplus W_1 \rightarrow W_1$ has the matrix

$$[\pi_1]_{\mathcal{B}_1} = \left[\begin{array}{c|c} 0 & P \\ \hline 0 & I_2 \end{array} \right].$$

Similarly, relative to \mathcal{B}_2 , a general complement Y_1 of Z is determined by a 3×1 matrix

$$\begin{bmatrix} I_1 \\ Q \end{bmatrix},$$

where

$$Q = \begin{bmatrix} -\gamma \\ -\delta \end{bmatrix}$$

is arbitrary, and the projection $\pi_2 : Y_1 \oplus Z \rightarrow Z$ has the matrix

$$[\pi_2]_{\mathcal{B}_2} = \left[\begin{array}{c|c} 0 & 0 \\ \hline -Q & I_2 \end{array} \right].$$

(We have introduced negatives in the entries of Q in order to avoid them in the matrix of the projection!) Notice that, in order to apply Lemma 4.5.1 to calculate $[\pi_2]_{\mathcal{B}_2}$, we needed to reorder the basis \mathcal{B}_2 so that the vectors spanning Z come first (which amounts to conjugation of the earlier projection matrix by the permutation matrix of the cyclic permutation (213)). Now by Proposition 4.4.3 the “most general” generalized inverse of a is

$$c = \pi_1 b \pi_2.$$

It remains only to compute the matrix of c in the standard basis \mathcal{B} . One quickly computes the following change of basis matrices:

$$S = [\mathcal{B}_1, \mathcal{B}] = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad S^{-1} = [\mathcal{B}, \mathcal{B}_1] = \frac{1}{2} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$T = [\mathcal{B}_2, \mathcal{B}] = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 3 \\ 0 & -2 & 2 \end{bmatrix}, \quad T^{-1} = [\mathcal{B}, \mathcal{B}_2] = \frac{1}{10} \begin{bmatrix} 10 & -4 & 6 \\ 0 & 2 & -3 \\ 0 & 2 & 2 \end{bmatrix}.$$

Now we have a description of all the generalized inverses C of our original matrix A :

$$\begin{aligned}
 C &= [c]_{\mathcal{B}} \\
 &= [\pi_1]_{\mathcal{B}} [b]_{\mathcal{B}} [\pi_2]_{\mathcal{B}} \\
 &= S [\pi_1]_{\mathcal{B}_1} S^{-1} [b]_{\mathcal{B}} T [\pi_2]_{\mathcal{B}_2} T^{-1} \\
 &= \begin{bmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & \alpha & \beta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{5} & -\frac{3}{10} \\ 0 & \frac{1}{5} & \frac{1}{5} \end{bmatrix} \\
 &\quad \cdot \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 3 \\ 0 & -2 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 \\ \gamma & 1 & 0 \\ \delta & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -\frac{2}{5} & \frac{3}{5} \\ 0 & \frac{1}{5} & -\frac{3}{10} \\ 0 & \frac{1}{5} & \frac{1}{5} \end{bmatrix} \\
 &= \frac{1}{10} \begin{bmatrix} 20\alpha\gamma + 20\beta\delta, & 4\alpha + 4\beta - 8\alpha\gamma - 8\beta\delta, & -6\alpha + 4\beta + 12\alpha\gamma + 12\beta\delta \\ 10\gamma - 10\alpha\gamma - 10\beta\delta, & 2 - 2\alpha - 2\beta - 4\gamma + 4\alpha\gamma + 4\beta\delta, & -3 + 3\alpha - 2\beta + 6\gamma - 6\alpha\gamma - 6\beta\delta \\ 10\delta, & 2 - 4\delta, & 2 + 6\delta \end{bmatrix}.
 \end{aligned}$$

By our construction, different choices of the four parameters $\alpha, \beta, \gamma, \delta$ give different generalized inverses of our matrix A (so we have managed to faithfully parameterize all the generalized inverses).¹⁹ \square

Example 4.5.4

In the case of $F = \mathbb{R}$ or $F = \mathbb{C}$, if we regard $V = F^n$ as an inner product space in the usual way, then the most natural choice for complementary subspaces X and Y in our first Example 4.4.2 are the orthogonal complements of $\ker(a)$ and $\text{im}(a)$, respectively. The uniquely determined generalized inverse of a in this case is the

19. For moderately larger $n \times n$ matrices A of nullity m , although it is not too difficult to parameterize the generalized inverses C of A when written as a product of seven matrices, explicitly describing the entries of such C in terms of the $m(n-m) + (n-m)m = 2(mn - m^2)$ parameters becomes very messy (even for $n = 4, m = 2$).

Moore–Penrose inverse²⁰ of a , which is denoted by a^+ . It is worth redrawing our earlier diagram to emphasize the connections, because the Moore–Penrose inverse is not always presented this way:

$$V = \left\{ \begin{array}{ccc} \ker(a) & \begin{array}{c} \xrightarrow{0} \\ \xleftarrow{0} \end{array} & \operatorname{im}(a)^\perp \\ \oplus & & \oplus \\ \ker(a)^\perp & \begin{array}{c} \xrightarrow{a} \\ \xleftarrow{a^+} \end{array} & \operatorname{im}(a) \end{array} \right.$$

where the symbol $^\perp$ indicates the orthogonal complement. Again, the forward arrows indicate the action of a , while the backward ones indicate the action of a^+ , which is the inverse map of a on the lower summands.

A more common way of presenting the Moore–Penrose inverse of a not necessarily square complex $m \times n$ matrix A is that it is characterized as the $n \times m$ matrix A^+ satisfying the following four conditions, where X^* denotes the conjugate transpose of a matrix X :

$$\begin{array}{lll} \text{(i)} & A & = AA^+A \\ \text{(ii)} & A^+ & = A^+AA^+ \\ \text{(iii)} & (AA^+)^* & = AA^+ \\ \text{(iv)} & (A^+A)^* & = A^+A \end{array}$$

From these equations, one quickly sees that the Moore–Penrose inverse of a Jordan matrix or a Weyr matrix is just its transpose. In general, the Moore–Penrose inverse can be explicitly calculated, in a number of ways. For instance, if one knows the singular value decomposition $A = Q_1 \Sigma Q_2^*$, then $A^+ = Q_2 \Sigma^+ Q_1^*$ (the middle term is trivial to compute, namely invert the nonzero entries on the main diagonal). See Strang's *Linear Algebra and Its Applications*, Section 6.3. The Moore–Penrose inverse can also be calculated by elementary row operations. If $\operatorname{rank} A = r$, then by using row operations we can obtain a *full rank*

20. E. I. Fredholm was the first to introduce the concept of a generalized inverse, for an integral operator, in a 1903 paper. However, generalized inverses for matrices did not appear until E. H. Moore's 1920 abstract in the *Bulletin of the American Mathematical Society*. Details of his work later appeared, posthumously, in his book *General Analysis, Part I*. This seems to have gone unnoticed until there was renewed interest in the early 1950s stemming from least squares problems. Then, while still a student (but not yet a knight), R. A. Penrose published a paper in 1955 showing that Moore's generalized inverse A^+ of the $m \times n$ complex matrix A is the unique matrix satisfying conditions (i)–(iv) listed above.

factorization

$$A = PQ$$

of A where P is $m \times r$, and Q is $r \times n$, and both factors have full (column or row) rank r . (For instance, if $CA = R$ gives a row echelon form of A , where C is invertible, throw away the last $m - r$ columns of C^{-1} to get P , and throw away the last $m - r$ rows of R to get Q .) Now

$$A^+ = Q^+ P^+.$$

But the Moore–Penrose inverse of a full rank matrix is known explicitly:

$$Q^+ = Q^*(QQ^*)^{-1}, \quad P^+ = (P^*P)^{-1}P^*.$$

Again see Strang's Section 6.3, Problems 19 and 22. Note that since the rank of a matrix X agrees with the ranks of XX^* and X^*X , the matrices QQ^* and P^*P are indeed invertible $r \times r$ matrices. \square

Remark. There is another way in which one would be naturally led to the Moore–Penrose inverse of an $n \times n$ matrix over \mathbb{C} . To fit with the discussion of quasi-inverses of elements of a general ring R , let us for just this once denote matrices in $M_n(\mathbb{C})$ by lower case. Let $R = M_n(\mathbb{C})$ and fix $a \in R$. Every left or right principal ideal of R is generated by a unique **projection**, that is, an idempotent e such that $e = e^*$ (self-adjoint). Write

$$Ra = Re, \quad aR = fR,$$

where e and f are projections. Now one returns to an earlier observation in Remark 4.4.5 (2) and asks what is the unique generalized inverse of a associated with this pair of idempotent generators? You guessed it—the Moore–Penrose inverse a^+ . \square

Example 4.5.5

Let's compute the Moore–Penrose inverse of the real matrix

$$A = \begin{bmatrix} 1 & 1 & -2 & 1 \\ 2 & 2 & -6 & 4 \\ 3 & 3 & -2 & -1 \\ -1 & -1 & 1 & 0 \end{bmatrix}.$$

Using elementary row operations, we find (for instance) that left multiplying A by the invertible matrix

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & -\frac{1}{2} & 0 & 0 \\ -7 & 2 & 1 & 0 \\ 2 & -\frac{1}{2} & 0 & 1 \end{bmatrix}$$

puts A in echelon form

$$CA = R = \begin{bmatrix} 1 & 1 & -2 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and also reveals $\text{rank } A = 2$. Now

$$A = C^{-1}R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & -2 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -2 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 2 & -2 \\ 3 & 4 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -2 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

$$= PQ, \text{ say.}$$

Since $\text{rank } A = 2$, the factorization $A = PQ$ is a full rank factorization. We can now use the formula in Example 4.5.4 to compute the Moore–Penrose inverse A^+ . We have

$$P^*P = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 0 & -2 & 4 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & -2 \\ 3 & 4 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 15 & 9 \\ 9 & 21 \end{bmatrix}$$

and

$$QQ^* = \begin{bmatrix} 1 & 1 & -2 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ -2 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 7 & -3 \\ -3 & 2 \end{bmatrix}.$$

Hence

$$A^+ = Q^*(QQ^*)^{-1}(P^*P)^{-1}P^*$$

$$\begin{aligned} &= \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ -2 & 1 \\ 1 & -1 \end{bmatrix} \frac{1}{5} \begin{bmatrix} 2 & 3 \\ 3 & 7 \end{bmatrix} \frac{1}{78} \begin{bmatrix} 7 & -3 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & -1 \\ 0 & -2 & 4 & -1 \end{bmatrix} \\ &= \frac{1}{390} \begin{bmatrix} 5 & -8 & 51 & -14 \\ 5 & -8 & 51 & -14 \\ -10 & -36 & 2 & 2 \\ 5 & 44 & -53 & 12 \end{bmatrix}. \end{aligned}$$

It's magic. □

4.6 THE JORDAN FORM DERIVED MODULE-THEORETICALLY

This section gives details of how the Jordan form can be deduced quite quickly from a well-known result on the decomposition of finitely generated modules M over a principal ideal domain R : such a module M is a direct sum of a finitely generated free R -module and a finitely generated torsion module. In turn, the torsion part decomposes uniquely (to within isomorphism) as a direct sum of cyclic modules whose annihilator ideals are principal ideals Rp^m generated by powers of irreducible (equivalently prime) elements $p \in R$. In actual fact, we need the result only in the case M is a finitely generated torsion module and R is the domain $F[x]$ of polynomials over an algebraically closed field. For those readers unfamiliar with the latter decomposition, one can draw the comparison in the case of a finite abelian group M regarded as a \mathbb{Z} -module: M is uniquely a direct sum of cyclic groups of prime power order.

We remind the reader of some earlier terminology. Let R be an integral domain (commutative ring without zero divisors) and M be an R -module.

Given a nonempty subset T of M , recall that its **annihilator ideal** $\text{ann}_R(T)$ is defined as

$$\text{ann}_R(T) = \{r \in R : rt = 0 \text{ for all } t \in T\}.$$

Since there should be no confusion, we simplify this notation to $\text{ann}(T)$. It is straightforward to see that $\text{ann}(T)$ is an ideal of R . If T is the singleton $\{t\}$, we write $\text{ann}(t)$ instead of $\text{ann}(T)$. If $\text{ann}(t) \neq 0$, we say that t is a **torsion element**. If each $t \in M$ is torsion, M is called a **torsion module**. At the other extreme, if 0 is the only torsion element in M , then M is called **torsion-free**. It's easily checked that every free R -module is torsion-free.²¹ If M is finitely generated by torsion elements t_1, t_2, \dots, t_k , then M is a torsion module for, if $0 \neq r_i \in \text{ann}(t_i)$ for each i , then $0 \neq r_1 r_2 \cdots r_k \in \text{ann}(M)$ (recalling that R is an integral domain). Cyclic modules are closely associated with annihilators for if $M = Ra$ then, applying the Fundamental Homomorphism Theorem 4.1.9 to the mapping $R \rightarrow M, r \mapsto ra$ gives $M \cong R/\text{ann}(a)$.

A **principal ideal domain** (PID) is an integral domain R for which each ideal takes the form $Ra = \{ra : r \in R\}$. The classic examples of a PID are the ring \mathbb{Z} of integers and the polynomial ring $F[x]$ over a field F . More generally, any Euclidean domain (one for which there is a division algorithm) is a PID.

Now let F be an algebraically closed field and let $R = F[x]$ be the ring of polynomials in the indeterminate x . Next, let $t : V \rightarrow V$ be a given linear transformation of some finite-dimensional vector space V over F . Then, as described in Example 4.1.7, we can convert V into an R -module with module multiplication given by

$$f(x) \cdot v = f(t)(v) \quad \text{for all } f(x) \in R, v \in V.$$

This module action, of course, depends on the fixed t but there shouldn't be confusion if we omit this dependence in the module notation V . Notice that the R -submodules of V are precisely the t -invariant subspaces. We present two critical lemmas.

Lemma 4.6.1

Let $R = F[x]$, V the R -module as above, and let U be an R -submodule of V . Then:

- (i) V is a finitely generated torsion R -module.
- (ii) $\text{ann}(U)$ is the principal ideal of R generated by the minimal polynomial $m(x)$ of the transformation t restricted to U .

21. The converse fails. For instance, \mathbb{Q} is a torsion-free \mathbb{Z} -module, but not a free \mathbb{Z} -module because it is indecomposable (any two nonzero subgroups of \mathbb{Q} have a nonzero intersection) but not cyclic.

- (iii) If $U = Ru$ is a cyclic R -submodule generated by u , and the minimal polynomial $m(x)$ in (ii) has degree n , then $\mathcal{B} = \{u, t(u), t^2(u), \dots, t^{n-1}(u)\}$ is a vector space basis of U over F .

Proof

(i) V is finitely generated over F , so certainly finitely generated over $R = F[x]$. Since the minimal polynomial of t annihilates all elements of V , we have that V is a torsion R -module.

(ii) Let $s = t|_U$. Then $m(t)(U) = m(s)(U) = 0$ so certainly $m \in \text{ann}(U)$ and hence $Rm \subseteq \text{ann}(U)$. On the other hand, if $f \in \text{ann}(U)$, then $f(s)(U) = f(t)(U) = 0$ and therefore $m(x)$ divides $f(x)$. Hence, $\text{ann}(U) \subseteq Rm$. Therefore, $\text{ann}(U) = Rm$.

(iii) Suppose $U = Ru$. Let $v \in U$, say $v = f(x) \cdot u = f(t)(u)$ where $f(x) \in R$. By the division algorithm, we can write $f(x) = q(x)m(x) + r(x)$ for some $q, r \in F[x]$ where either $r = 0$ or $\deg(r) < n$. Now

$$v = f(t)(u) = q(t)(u)m(t)(u) + r(t)(u) = 0 + r(t)(u) = r(t)(u),$$

which shows v is in the span of \mathcal{B} . Therefore, \mathcal{B} spans U . The members of \mathcal{B} are also linearly independent, because if $\lambda_0 u + \lambda_1 t(u) + \dots + \lambda_{n-1} t^{n-1}(u) = 0$ for $\lambda_i \in F$, then the polynomial $f(x) = \lambda_0 + \lambda_1 x + \dots + \lambda_{n-1} x^{n-1}$ is in $\text{ann}(U)$, whence by (ii) $f(x)$ is divisible by $m(x)$. By degree comparisons, this forces $f = 0$ and therefore all $\lambda_i = 0$. Hence, \mathcal{B} is a basis for U . \square

Lemma 4.6.2

In the same notation as in Lemma 4.6.1, suppose U is a cyclic R -submodule of V with $\text{ann}(U) = R(x - \lambda)^n$ for some scalar $\lambda \in F$ and positive integer n . Then there exists a basis \mathcal{B} for U such that the matrix $J = [t|_U]_{\mathcal{B}}$ of the transformation t restricted to U is the $n \times n$ basic Jordan matrix

$$J = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{bmatrix}.$$

Proof

It is enough to do the case $\lambda = 0$ (t nilpotent), because for a general λ we can replace t by $t - \lambda 1$ and use the fact that

$$[t|_U]_{\mathcal{B}} = [(t - \lambda 1)|_U]_{\mathcal{B}} + \lambda I.$$

Notice that under the new module action $f(x)(v) = f(t - \lambda 1)(v)$, the subspace U is still cyclic but now $\text{ann}(U) = Rx^n$. Thus, $(t - \lambda 1)|_U$ is nilpotent of index n by Lemma 4.6.1 (ii).

So assume $\lambda = 0$ and $U = Ru$ for some $u \in U$. Let $s = t|_U$. The minimal polynomial of s is x^n and so by Lemma 4.6.1 (iii) we know $\{u, t(u), t^2(u), \dots, t^{n-1}(u)\}$ is a basis for U . Because we are after an upper triangular Jordan block, we need to reorder this basis, by running through the vectors in reverse order. Thus, we choose the basis

$$\mathcal{B} = \{t^{n-1}(u), t^{n-2}(u), \dots, t^2(u), t(u), u\} = \{v_1, v_2, \dots, v_n\}.$$

The action of s on the basis vectors is now the standard shift, shift, \dots , annihilate (moving from right to left):

$$\begin{aligned} s(v_1) &= 0 \\ s(v_2) &= v_1 \\ s(v_3) &= v_2 \\ &\vdots \\ s(v_n) &= v_{n-1} \end{aligned}$$

Hence,

$$[s]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix},$$

which is the $n \times n$ basic nilpotent Jordan matrix. We are done. \square

Recall that an element p of an integral domain R is **irreducible** if p is not a unit (i.e., not invertible) and whenever $p = ab$ where $a, b \in R$, then either a or b is a unit (so p has only the trivial factorizations). Every PID is a **unique factorization domain** (UFD), that is every nonzero, nonunit element has a factorization into irreducibles, which is unique to within the order of the factors and unit multiples. In particular $F[x]$ is a UFD.²² When F is algebraically closed, what is special about the polynomials $(x - \lambda)^n$ in Lemma 4.6.2 is that, to within scalar multiples, they are exactly the powers of irreducible polynomials in $F[x]$.

22. The polynomial ring $F[x_1, x_2, \dots, x_n]$ in two or more indeterminates is also a UFD, but not a PID. More generally, if R is a UFD, then so is the polynomial ring $R[x]$. See Jacobson's *Basic Algebra I*, Chapter 2.

It's now time to roll in the big gun.

Their definition leaves no doubt that PID's have a very special internal structure. It should come as no surprise then that these domains have also a well-determined module theory. For example, the PID's are precisely the commutative rings R for which the submodules of every free R -module are again free.²³ Moreover, if R is a PID then R has the **FGC property**, namely any finitely generated R -module splits into a direct sum of cyclic submodules.²⁴ We only need the torsion part of this, which we record in the following theorem. We omit its proof—the hungry reader can find it in several standard texts, for example, Jacobson's *Basic Algebra I*, Section 3.8. The Hartley and Hawkes text *Rings, Modules and Linear Algebra* also has an excellent treatment of the theorem and its corollaries.

Theorem 4.6.3 (The Fundamental Theorem for Finitely Generated Modules over a PID: Torsion Case)

Let M be a nonzero, finitely generated torsion module over a principal ideal domain R . Then there is a direct sum decomposition

$$M = U_1 \oplus U_2 \oplus \cdots \oplus U_k$$

of M into cyclic submodules U_i whose annihilator ideals $\text{ann}(U_i)$ are the principal ideals $Rp_i^{m_i}$ of R generated by powers of irreducible elements p_i . Moreover, this decomposition is unique to within isomorphism and the order of the summands.

Note: The irreducibles p_1, p_2, \dots, p_k are not necessarily distinct. But they are uniquely determined to within order and unit multiples. □

All that remains to establish the Jordan form is to load Lemmas 4.6.1 and 4.6.2 into our gun, Theorem 4.6.3, and fire.

Derivation of the Jordan form. Let $A \in M_n(F)$, where F is an algebraically closed field. Let $R = F[x]$. Let $V = F^n$ be the space of all $n \times 1$ column vectors and let $t : V \rightarrow V$ be the linear transformation given by left multiplication by A . Endow V with the structure of an R -module using t , as earlier described in Example 4.1.7. By Lemma 4.6.1, V is a finitely generated, torsion R -module over the principal ideal domain R . Hence, by the Fundamental Theorem 4.6.3, there

23. See Corollary 6.4 of D. Passman's text *A Course in Ring Theory* for a proof of the more difficult part of this.

24. PID's are not the only commutative rings, or indeed the only integral domains, with the FGC property. These rings were not completely classified until 1976. Details can be found in W. Brandal's 1979 monograph *Commutative Rings Whose Finitely Generated Modules Decompose*. The decomposition of all finitely generated torsion modules into cyclics is closely linked to the prime ideal structure of the underlying ring, as can be seen in a 1990 article by Clark (our second author), Brandal, and Barbut.

is a decomposition

$$V = U_1 \oplus U_2 \oplus \cdots \oplus U_k$$

of V in which the U_i are cyclic submodules whose annihilator ideals are generated by $(x - \lambda_i)^{n_i}$ for some scalars λ_i and positive integers n_i . By Lemma 4.6.2, there is a basis \mathcal{B}_i for U_i such that the matrix J_i of the restriction of t to U_i is

$$J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i & 1 \\ & & & & & \lambda_i \end{bmatrix}.$$

Let $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$. This is a basis for V and the matrix $J = [t]_{\mathcal{B}}$ of t relative to \mathcal{B} is block diagonal with the J_i as its diagonal blocks because the U_i , being R -submodules of V , are invariant under t . See Proposition 1.3.2. By reordering the blocks of J appropriately (by a permutation conjugation), we have J in Jordan form. But A is similar to J (two matrices of t in different bases). Therefore, A has a Jordan form.

The uniqueness of the Jordan form, which we established in Corollary 2.4.6, can also be deduced by the uniqueness part of the fundamental theorem. \square

We've crossed the Jordan. Now on to the Promised Land.

4.7 THE WEYR FORM OF A NILPOTENT ENDOMORPHISM: PHILOSOPHY

Here we discuss the possibility of formulating a Weyr form of a general nilpotent endomorphism $t : P \rightarrow P$ of *any* quasi-projective module P over *any* ring R , and without finiteness conditions on P . It does not appear to have a true Jordan analogue. The next section will then provide necessary and sufficient conditions for the existence of the form.

The quintessential nilpotent transformation $t : V \rightarrow V$ of an n -dimensional vector space V over a field F is the one whose matrix relative to some basis

$\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ is a basic nilpotent Jordan matrix

$$\begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{bmatrix}.$$

Directly, in terms of t , this means t annihilates v_1 and then shifts in order each of the other v_i to its immediate predecessor v_{i-1} :

$$0 \longleftarrow v_1 \longleftarrow v_2 \longleftarrow \cdots \longleftarrow v_{n-1} \longleftarrow v_n.$$

To within scalar multiples of the basis elements, choosing a basis amounts really to just specifying a direct sum decomposition

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_n$$

of V into 1-dimensional subspaces. The action of our quintessential t on these subspaces in the case of the good basis \mathcal{B} is the same shifting and annihilating as above if we replace the vector v_i by the subspace V_i . If we were not given the basis \mathcal{B} but instead the direct sum decomposition with the above shifting action of t on the 1-dimensional summands, we could recover a suitable basis \mathcal{B} that would yield the Jordan block, by choosing any nonzero v_n in V_n and then recursively taking $v_{i-1} = t(v_i)$ for $i = n, \dots, 2$. This “*shift, shift, \dots, shift, annihilate*” view (moving from right to left) is so natural. Does it usefully generalize? It turns out that it does, as we will show, but the extension is suggested much more by the Weyr form than the Jordan form. Note that from a module point of view, for a general nilpotent transformation t , if its matrix relative to some basis is in Jordan form with Jordan structure (m_1, m_2, \dots, m_s) , this corresponds to a direct sum decomposition $V = U_1 \oplus U_2 \oplus \cdots \oplus U_s$ of V into nonzero t -invariant subspaces U_i of dimension m_i such that t acts on each U_i as the quintessential nilpotent 1-dimensional shift transformation.

What is the Weyr form saying about a linear transformation $t : V \rightarrow V$? In the nilpotent case, if the basis $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ gives the matrix of t in Weyr form with Weyr structure (n_1, n_2, \dots, n_r) , we can form a direct sum decomposition $V = V_1 \oplus V_2 \oplus \cdots \oplus V_r$ of V where V_1 is spanned by the first n_1 basis vectors, V_2 by the next n_2 basis vectors, and so on down to V_r being spanned by the last n_r basis vectors. And the action of t is to annihilate V_1 and then map

V_i into V_{i-1} for $i = 2, 3, \dots, r$ by shifting, in order, the aforementioned n_i basis vectors in V_i to the corresponding first n_i of the n_{i-1} basis vectors of V_{i-1} . In particular, t maps V_i *isomorphically* onto a subspace of V_{i-1} for $i = 2, \dots, r$. This is the critical feature (along with t also annihilating V_1). The Weyr form incorporates the natural “*shift, shift, \dots, shift, annihilate*” phenomenon for *all* nilpotent transformations (unlike the Jordan form). If we didn’t know the basis \mathcal{B} but knew a direct sum decomposition $V = V_1 \oplus V_2 \oplus \dots \oplus V_r$ of V relative to which t exhibits this shifting feature, we could recover a suitable basis to provide the Weyr matrix form of t . Namely, we could start with any basis \mathcal{B}_r for V_r . Then we could extend $t(\mathcal{B}_r)$ to a basis for V_{r-1} and call this \mathcal{B}_{r-1} . Next we could extend $t(\mathcal{B}_{r-1})$ to a basis for V_{r-2} and call this \mathcal{B}_{r-2} . We would continue in this way to recursively construct the \mathcal{B}_i for $i = r-1, r-2, \dots, 1$. Finally we would set $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_r$.

The reader may be puzzled as to why our shifts are from right to left, not left to right. It simply comes down to whether one has a preference for *upper* triangular canonical forms or *lower* triangular ones. We have gone for upper triangular, which is probably the more popular choice worldwide. However, it is a bit like a nation driving on the left side or the right side of the road—the latter is more common, but far from universal.²⁵

What are the possibilities (and difficulties) for extending the above decomposition ideas associated with a nilpotent linear transformation to a more general nilpotent endomorphism $t : M \rightarrow M$ of a module M over some general ring R (not even commutative). At the very least, we should insist that any generalized Jordan or Weyr decompositions, when specialized to a linear transformation of a finite-dimensional vector space, should agree with the classical Jordan (Weyr) decompositions that we have just been discussing. Since fields F are such special rings, modules over them ($R = F$) have almost any nice property you care to name. All vector spaces are free F -modules, in particular projective. Working with a projective module M over a ring R provides a possibly more general setting. However, here the Jordan form analysis still presents immediate problems because of the very special nature of 1-dimensional subspaces used in the Jordan description given in Chapter 1, Section 1.6.

Among the nonzero F -submodules of a vector space V , the 1-dimensional submodules can be characterized in various module-theoretic ways: simple, or indecomposable, or cyclic. These are all equivalent for subspaces. For a general module M , a decomposition into a direct sum of simple submodules is usually a pretty special situation (the semisimple modules), but a decomposition into a direct sum of indecomposable submodules will hold under fairly mild

25. Due to the British influence, the first and second authors drive on the left.

“finiteness” conditions (for example, if there is a finite bound to the number of nonzero submodules occurring in any direct sum decomposition of M). But even a projective module may not possess a single indecomposable submodule!²⁶ To ensure a module M is a finite direct sum of cyclic modules, one firstly needs M to be finitely generated, and if all such modules over R have this cyclic decomposition, i.e., R has the FGC property, then, as mentioned earlier, R will usually be fairly special, such as a principal ideal domain. Of course, with each of the three candidates (simple, indecomposable, cyclic) discussed as possible summands in some decomposition of M relative to t , there is still the daunting task of actually relating them to t in some revealing way. After all, the goal is to take t apart to see what makes it tick. Thus, generalizing the Jordan form from a module point of view is not too promising. We take up this topic again in Section 4.9.

With the Weyr form view of a linear transformation, the situation for generalizing is much more hopeful because, on the surface at least, there are no real constraints on the summands V_i —their dimensions are restricted only by being decreasing and summing to $\dim V$. In fact, our discussion above already suggests a formulation, albeit an ambitious one, to more general nilpotent module endomorphisms of even a quasi-projective module, as in the following definition. When restricted to a linear transformation, this definition will agree with the one for the (unique) Weyr form (see proof of Corollary 4.8.4).

*Definition 4.7.1: Suppose $t : P \rightarrow P$ is a nilpotent endomorphism of a (nonzero) quasi-projective module P over an arbitrary (noncommutative)²⁷ ring R . Then a **Weyr form** for t is a direct sum decomposition*

$$P = P_1 \oplus P_2 \oplus \cdots \oplus P_r$$

of P into nonzero submodules such that t annihilates P_1 and maps P_i isomorphically onto a direct summand of P_{i-1} for $i = 2, 3, \dots, r$.

In the Definition 4.7.1, one could quibble over whether “direct summand of P_{i-1} ” should be weakened to “submodule of P_{i-1} .” In the vector space setting, it makes no difference (all subspaces are direct summands). For the general quasi-projective module case, we feel the stronger requirement is a more accurate reflection of the vector space situation.

26. One example of this is to take S to be the ring of all linear transformations of a countably-infinite dimensional vector space and let $R = S/I$ where I is the ideal of all transformations of finite rank. Then R is a (simple) von Neumann regular ring with $Ra \cong Ra \oplus Ra$ for all nonzero $a \in R$. It follows that R has no indecomposable left ideals, whence R as a left R -module has no indecomposable submodules.

27. Here, and elsewhere, “noncommutative” means “not necessarily commutative.”

Of course, it is one thing to make a definition; it is another to show its usefulness. In Theorem 4.8.2, we give a precise (and verifiable) condition for t (as in the definition) to actually have a Weyr decomposition, viz. all the powers of t must be regular in the endomorphism ring $\text{End}_R(M)$. Thus, the “philosophizing” of this section will be followed up with the “nitty-gritty” in the next section.

4.8 THE WEYR FORM OF A NILPOTENT ENDOMORPHISM: EXISTENCE

In this section, we establish the conditions for the existence of a Weyr form of a nilpotent endomorphism $t : P \rightarrow P$ of a quasi-projective module P , in two steps. The first step, Proposition 4.8.1, is the critical one. It is only a slight modification of Lemma 7.1 in Ken Goodearl’s book *Von Neumann Regular Rings*, concerning a nilpotent endomorphism $t : P \rightarrow P$ of a finitely generated projective module P over a regular ring R . Goodearl gives a direct sum decomposition like that in Definition 4.7.1 except that after annihilating P_1 , t maps each P_i onto P_{i-1} , not necessarily isomorphically, for $i = 2, 3, \dots, r$. (Definition 4.7.1 on the other hand requires t to map P_i isomorphically onto a direct summand of P_{i-1} .) We would like to express our appreciation to Ken Goodearl for pointing out the possible connection of his lemma with the Weyr form, during a talk by the first author at Santa Barbara in 2004.

In the Goodearl version, the ring $\text{End}_R(P)$ of R -endomorphisms of P is known to be a regular ring. Using essentially the same ideas as Goodearl (but with a different induction, and a different indexing because we want the Weyr form to be upper triangular, not lower), we reach the same conclusions as he did, but under weaker hypotheses. We drop the regularity requirement for the ring R and the finitely generated projective requirement for the module P , and in their place insert the very much weaker requirement that t and all its powers be regular in $\text{End}_R(P)$, where P is now any quasi-projective R -module over an arbitrary ring R .²⁸

Proposition 4.8.1 (**Goodearl**)

Let R be any ring, let P be a nonzero quasi-projective module over R , and let $t : P \rightarrow P$ be a nilpotent R -endomorphism of P of index r . Assume that t and all its powers are regular in $\text{End}_R(P)$. Then there is a decomposition

$$P = A_1 \oplus A_2 \oplus \cdots \oplus A_r$$

28. This modified version was first used by Beidar, O’Meara, and Raphael in another setting in 2004.

of P into r nonzero submodules such that $t(A_1) = 0$ and $t(A_i) = A_{i-1}$ for $i = 2, \dots, r$.

Proof

We argue by induction on the nilpotent index r of t . If $r = 1$, then t is the zero mapping and we can take $A_1 = P$ to establish the proposition. Now assume $r > 1$ and that the result holds for nilpotent endomorphisms of index $r - 1$. Let $Y = t(P)$ and let $s : Y \rightarrow Y$ be the restriction of t to Y . Since t is a regular endomorphism of P , we know Y is a direct summand of P by Proposition 4.4.6. Hence, Y is a quasi-projective R -module by Proposition 4.3.7 (1). Given a positive integer k , our assumption about the regularity of powers of t implies that $t^{k+1}P$ is a direct summand of P , whence by Lemma 4.2.7 also a direct summand of Y . This just says that $s^k Y$ is a direct summand of Y , and therefore s^k is a regular endomorphism of Y by Proposition 4.4.6 because Y is quasi-projective. Thus, all powers of s are regular in $\text{End}_R(Y)$. Clearly, s is a nilpotent endomorphism of Y of index $r - 1$. Therefore, by our induction hypothesis, there are submodules A_1, A_2, \dots, A_{r-1} of Y such that

$$Y = A_1 \oplus A_2 \oplus \dots \oplus A_{r-1},$$

$s(A_1) = 0$, and $s(A_i) = A_{i-1}$ for $i = 2, 3, \dots, r - 1$. So to complete the proof, all we need do is construct a submodule A_r of P such that $P = Y \oplus A_r$ and $t(A_r) = A_{r-1}$.

We shall keep chopping P into pieces (direct sums), using several intermediate summands along the way, but with a clear goal in mind. The main summands to watch out for are indicated in the following diagram, where $P = Y \oplus D \oplus E$, $Y = t^2(P) \oplus A_{r-1}$, and the indicated maps are onto. We will then take $A_r = D \oplus E$.

$$P = \left\{ \begin{array}{ccc} Y & \xrightarrow{t} & t^2(P) \\ \oplus & & \oplus \\ D & \xrightarrow{t} & A_{r-1} \\ \oplus & & \oplus \\ E & \xrightarrow{t} & 0 \end{array} \right\} = Y$$

Notice that from the decomposition of Y and the action of t on the summands, we have

$$Y = t^2(P) \oplus A_{r-1}. \quad (1)$$

Set $B = t^{-1}(t^2P)$ and $C = t^{-1}(A_{r-1})$.²⁹ One quickly checks from (1) that $P = B + C$ because t maps P onto Y . Let $K = \ker(t)$. We know from Proposition 4.4.6 that K is a direct summand of P because t is a regular endomorphism of P . Since C is a submodule of P containing K , by Lemma 4.2.7 we obtain

$$C = D \oplus K \quad (2)$$

for some submodule D . Now $P = B + C = B + D + K = B + D$ because B contains K . Moreover $B \cap D \subseteq B \cap C \subseteq K$ because $t(B) \cap t(C) = t^2(P) \cap A_{r-1} = 0$. From $K \cap D = 0$, we therefore conclude that $B \cap D = 0$ and so now we have

$$P = B \oplus D. \quad (3)$$

We next establish that

$$B = Y \oplus E \quad (4)$$

for some submodule E of K . Observe that $t(Y) = t^2(P) = t(B)$, whence $Y \subseteq B$ and $B = Y + K$. As we earlier observed, t restricted to Y is a regular endomorphism, whence its kernel, $Y \cap K$, is a direct summand of Y and therefore a direct summand of K , say $K = (Y \cap K) \oplus E$ for some submodule E . Note $E \cap Y = 0$. Hence, $B = Y + K = Y + (Y \cap K) + E = Y \oplus E$, as desired.

Now to put Humpty-Dumpty back together again. Set $A_r = D + E = D \oplus E$, the sum being direct because $D \cap E \subseteq D \cap K = 0$. We have

$$\begin{aligned} P &= B \oplus D && \text{by (3)} \\ &= Y \oplus E \oplus D && \text{by (4)} \\ &= Y \oplus A_r && \text{by the definition of } A_r \end{aligned}$$

and

$$\begin{aligned} t(A_r) &= t(D + E) = t(D) + t(E) \\ &= t(D) && \text{because } E \subseteq K \\ &= t(D) + t(K) \\ &= t(D + K) = t(C) && \text{by (2)} \\ &= A_{r-1}. \end{aligned}$$

As argued in the opening paragraph, this is mission accomplished. \square

29. Here, $t^{-1}(X) = \{p \in P : t(p) \in X\}$ denotes the inverse image of a submodule X of P . It is also a submodule.

We are now ready to proceed to the second phase of our argument in establishing necessary and sufficient conditions for the existence of a Weyr form of a nilpotent endomorphism.

Theorem 4.8.2

Let R be any ring, P a nonzero quasi-projective module over R , and let $t : P \rightarrow P$ be a nilpotent R -endomorphism of P . Then t has a Weyr form if and only if all the powers of t are regular in the endomorphism ring $\text{End}_R(P)$.

Proof

First, assume the powers of t are regular. Let r be the nilpotent index of t . By Proposition 4.8.1 there is a direct sum decomposition $P = A_1 \oplus A_2 \oplus \cdots \oplus A_r$ such that $t(A_1) = 0$ and $t(A_i) = A_{i-1}$ for $i = 2, 3, \dots, r$. Set $B_{11} = A_1$. Since P is quasi-projective, so are its direct summands by Theorem 4.3.7 (1). In particular, the submodule $Q = A_1 \oplus A_2$ is quasi-projective. Therefore, by Proposition 4.3.8, the restriction of the mapping t to A_2 (which maps A_2 onto A_1) must split. Thus, we can write $A_2 = B_{12} \oplus B_{22}$ such that t maps B_{12} isomorphically onto B_{11} and $t(B_{22}) = 0$. Applying the same argument to t mapping A_3 onto A_2 , we obtain a decomposition $A_3 = B_{13} \oplus B_{23} \oplus B_{33}$ such that t maps B_{13} isomorphically onto B_{12} , B_{23} isomorphically onto B_{22} , and $t(B_{33}) = 0$. Continuing in this fashion, we get decompositions

$$A_j = \bigoplus_{i=1}^j B_{ij}$$

for $j = 1, 2, \dots, r$ such that $t(B_{ii}) = 0$ and t maps B_{ij} isomorphically onto $B_{i,j-1}$ for $j = i + 1, \dots, r$. Things become much clearer when we display the decompositions schematically:

$$\begin{array}{ccccccc}
 B_{11} & \leftarrow & B_{12} & \leftarrow & B_{13} & \leftarrow & \cdots & \leftarrow & B_{1r} \\
 & & \oplus & & \oplus & & & & \oplus \\
 & & B_{22} & \leftarrow & B_{23} & \leftarrow & \cdots & \leftarrow & B_{2r} \\
 & & & & \oplus & & & & \oplus \\
 & & & & B_{33} & \leftarrow & \cdots & \leftarrow & B_{3r} \\
 & & & & & & & & \oplus \\
 & & & & & & & & \vdots \\
 & & & & & & & & \oplus \\
 & & & & & & & & B_{rr}
 \end{array}$$

In this scheme,³⁰ the summands in the j th column sum to A_j , whence P is the direct sum of all the B_{ij} . The arrows represent isomorphic mappings under the restriction

30. We use the term “scheme” informally.

of t . The kernel of t is the sum $B_{11} \oplus B_{22} \oplus \cdots \oplus B_{rr}$ of the diagonal summands. Thus, in terms of the decomposition

$$P = \bigoplus_{1 \leq i \leq j \leq r} B_{ij},$$

the action of t in terms of faithful shifts and annihilations is very explicit. In fact, if the B_{ij} were something like “1-dimensional subspaces,” this is looking just like a Jordan decomposition! (See the comments at the beginning of Section 4.7, and the discussion to follow in Section 4.9.) To get the Weyr form, we proceed as follows.

For $i = 1, 2, \dots, r$, set

$$P_i = \bigoplus_{j=1}^{r-i+1} B_{j,j+i-1},$$

which is the sum of the summands in the i th diagonal of the above scheme. Thus,

$P_1 = B_{11} \oplus B_{22} \oplus B_{33} \oplus \cdots \oplus B_{rr}$	sum of 1st diagonal summands,
$P_2 = B_{12} \oplus B_{23} \oplus B_{34} \oplus \cdots \oplus B_{r-1,r}$	sum of 2nd diagonal summands,
$P_3 = B_{13} \oplus B_{24} \oplus B_{35} \oplus \cdots \oplus B_{r-2,r}$	sum of 3rd diagonal summands,
\vdots	\vdots
$P_r = B_{1r}$	sum of r th diagonal summands.

The above scheme now makes it clear that $P = P_1 \oplus P_2 \oplus \cdots \oplus P_r$ is a Weyr form for t . This is because $t(P_1) = 0$ and for $i = 2, 3, \dots, r$, we have $P_{i-1} = t(P_i) \oplus B_{r-i+2,r}$ whence t maps P_i isomorphically onto a direct summand of P_{i-1} . Note that the B_{1j} must be nonzero for $j = 1, \dots, r$ because t has nilpotent index r , and therefore all the P_i are nonzero.

For the converse, suppose $t : P \rightarrow P$ has a Weyr form $P = P_1 \oplus P_2 \oplus \cdots \oplus P_r$. Observe from the Weyr form that t has nilpotent index r and kernel P_1 . We wish to show t^k is a regular endomorphism of P for each $k = 1, 2, \dots, r-1$. By Proposition 4.4.6, since P is quasi-projective, it suffices to show $t^k(P)$ is a direct summand of P . This would be clear if we could produce a decomposition $P = A_1 \oplus A_2 \oplus \cdots \oplus A_r$ as in the beginning of the proof (i.e., as in Proposition 4.8.1). For then $t^k(P) = A_1 \oplus A_2 \oplus \cdots \oplus A_{r-k}$, which is a direct summand of P . We can indeed produce such a decomposition by reconstructing the earlier scheme

of B_{ij} for $i \leq j \leq r$ and then setting

$$A_j = \bigoplus_{i=1}^j B_{ij}.$$

These arguments are now becoming familiar (hopefully), so a quick sketch of the details will suffice. Unlike earlier, we produce the B_{ij} in the order they will appear within the diagonals, starting from the outermost r th diagonal and working inwards.

Set $B_{1r} = P_r$. Since t maps P_r isomorphically onto a direct summand of P_{r-1} , we can write $P_{r-1} = B_{1,r-1} \oplus B_{2r}$ where t maps B_{1r} isomorphically onto $B_{1,r-1}$ and B_{2r} is some submodule of P_{r-1} . Since t maps P_{r-1} isomorphically onto a direct summand of P_{r-2} , we can write $P_{r-2} = B_{1,r-2} \oplus B_{2,r-1} \oplus B_{3r}$ where t maps $B_{1,r-1}$ onto $B_{1,r-2}$, B_{2r} isomorphically onto $B_{2,r-1}$, and B_{3r} is some submodule of P_{r-2} . The pattern is clear. The final step, using the fact that t maps P_2 isomorphically onto a direct summand of P_1 , is to write $P_1 = B_{11} \oplus B_{22} \oplus \cdots \oplus B_{rr}$ where t maps $B_{i,i+1}$ isomorphically onto B_{ii} for $i = 1, 2, \dots, r-1$, and where B_{rr} is some submodule of P_1 . This completes our proof. \square

Remark 4.8.3

The scheme of the B_{ij} displayed in the proof of the theorem

$$\begin{array}{ccccccc}
 B_{11} & \leftarrow & B_{12} & \leftarrow & B_{13} & \leftarrow & \cdots & \leftarrow & B_{1r} \\
 & & \oplus & & \oplus & & & & \oplus \\
 & & B_{22} & \leftarrow & B_{23} & \leftarrow & \cdots & \leftarrow & B_{2r} \\
 & & & & \oplus & & & & \oplus \\
 & & & & B_{33} & \leftarrow & \cdots & \leftarrow & B_{3r} \\
 & & & & & & & & \oplus \\
 & & & & & & & & \vdots \\
 & & & & & & & & \oplus \\
 & & & & & & & & B_{rr}
 \end{array}$$

embodies three important decompositions of the quasi-projective module P , relative to the given nilpotent endomorphism t , in terms of the direct sum decompositions of P associated with the rows, columns, and diagonals, respectively. In the case of the column and diagonal decompositions, we lump together as a single summand the sum of all the B_{ij} within a given column or diagonal, respectively. There is a natural induced action (from right to left) of t on the new summands in

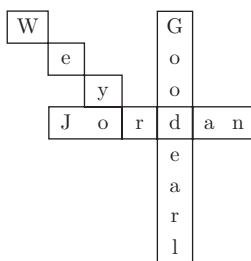


Figure 4.1 Three wise men, showing the way (moving to the left).

each of the three cases. Now:

The rows give a Jordan decomposition (4.9.1).

The columns give a Goodearl decomposition (4.8.1).

The diagonals give a Weyr decomposition (4.7.1).

However, for a nilpotent linear transformation $t : P \rightarrow P$ of a finite-dimensional vector space P , although the Weyr and Goodearl decompositions are unique (to within isomorphism), a Jordan decomposition in the above scheme won't correspond to "the" Jordan decomposition unless its summands are 1-dimensional. To help the reader remember which decomposition is which, we provide the little Figure 4.1 above. It should be carried at all times. \square

As a corollary to Theorem 4.8.2, we have our third independent way of verifying the existence of the Weyr form of a matrix.

Corollary 4.8.4

Every $n \times n$ matrix A over an algebraically closed field F is similar to a matrix in Weyr form.

Proof

By the standard reduction, we can assume A is a nilpotent matrix, say of index r . Let P be the space F^n of all $n \times 1$ column vectors and let $t : P \rightarrow P$ be the linear transformation given by left multiplication by A . Regarding P as a module over the ring $R = F$, certainly P is quasi-projective (in fact, free) and t is a nilpotent endomorphism of P of index r . Also the powers of t are regular by Example 4.4.2. Hence, Theorem 4.8.2 applies and so t has a Weyr form decomposition as a nilpotent endomorphism, say $P = P_1 \oplus P_2 \oplus \cdots \oplus P_r$. This, as we observed in Section 4.7, leads to a basis \mathcal{B} relative to which the matrix of t , say W , is in Weyr form. Namely, we start by choosing any basis \mathcal{B}_r for P_r . Since the restriction of t to P_r is a one-to-one linear transformation into P_{r-1} , we can extend $t(\mathcal{B}_r)$ to a basis for P_{r-1} , which we call \mathcal{B}_{r-1} . Next we extend

$t(\mathcal{B}_{r-1})$ to a basis for P_{r-2} , which we call \mathcal{B}_{r-2} . We continue in this way to recursively construct bases \mathcal{B}_i for each P_i for $i = r, r-1, \dots, 1$. Finally we set $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_r$. A quick check confirms W is in Weyr form with Weyr structure (n_1, n_2, \dots, n_r) , where $n_i = \dim P_i$. Of course, A is similar to W , so we are done. \square

4.9 A SMALLER UNIVERSE FOR THE JORDAN FORM?

Where Weyr goes, Jordan goes too, surely because of duality? Well, it seems not. Our formulation in Sections 4.7 and 4.8 of a Weyr form of a nilpotent endomorphism $t : P \rightarrow P$ of a quasi-projective module P does specialize uniquely in the case of a linear transformation of a finite-dimensional vector space. But as we show in this section, such a formulation for the Jordan form is not available without further restrictions.

The Weyr form of a nilpotent linear transformation $t : P \rightarrow P$ of a finite-dimensional vector space is unique in the strongest possible way: given two Weyr decompositions of P for t ,

$$P = P_1 \oplus P_2 \oplus \dots \oplus P_r$$

$$P = P'_1 \oplus P'_2 \oplus \dots \oplus P'_s$$

we must have $r = s$ and $\dim P_i = \dim P'_i$ for $i = 1, \dots, r$ (thus, there is a vector space automorphism of P that maps the first decomposition to the second). We can deduce this from the uniqueness of the Weyr form of a matrix for t (Proposition 2.2.3) because, as we saw in the proof of Corollary 4.8.4, $(\dim P_1, \dim P_2, \dots, \dim P_r)$ and $(\dim P'_1, \dim P'_2, \dots, \dim P'_s)$ each gives the Weyr structure of the matrix of t .

In the broad setting of Definition 4.7.1, there is no clear natural formulation of a “Jordan form” for a nilpotent endomorphism $t : P \rightarrow P$ of a quasi-projective module P , given one wants uniqueness in the case of a linear transformation. From our discussion at the beginning of Section 4.7, one reasonable definition of a “Jordan form” for t might be as follows.

*Definition 4.9.1: Suppose $t : P \rightarrow P$ is a nilpotent endomorphism of a nonzero quasi-projective module P over an arbitrary (noncommutative) ring R . Then a **Jordan form** for t is a direct sum decomposition*

$$P = M_1 \oplus M_2 \oplus \dots \oplus M_s$$

of P into nonzero t -invariant submodules M_i such that in turn, each M_i decomposes as

$$M_i = N_{i1} \oplus N_{i2} \oplus \dots \oplus N_{im_i}$$

where t annihilates N_{i1} and maps N_{ij} isomorphically onto $N_{i,j-1}$ for $j = 2, \dots, m_i$.

The proof of Theorem 4.8.2 shows that a Jordan decomposition in this sense exists if (and only if) the powers of t are regular. In the notation used there, we can take $M_i = B_{ii} \oplus B_{i,i+1} \oplus \dots \oplus B_{ir}$ for those indices i for which B_{ir} is nonzero, and let $N_{ij} = B_{i,i+j-1}$ for $j = 1, \dots, r - i + 1$. The problem with Definition 4.9.1, however, is that even a nilpotent linear transformation of a finite-dimensional vector space will have many quite different Jordan forms. For instance, consider the nilpotent linear transformation t on $P = F^6$ whose matrix J relative to some basis $\{v_1, v_2, v_3, v_4, v_5, v_6\}$ is the sum of three 2×2 Jordan blocks:

$$J = \left[\begin{array}{cc|cc|cc} 0 & 1 & & & & \\ 0 & 0 & & & & \\ \hline & & 0 & 1 & & \\ & & 0 & 0 & & \\ \hline & & & & 0 & 1 \\ & & & & 0 & 0 \end{array} \right].$$

One can see that there are (at least) three “nonisomorphic” decompositions of P meeting the Jordan form requirements in Definition 4.9.1:

$$P = M_1 \oplus M_2 \oplus M_3 \text{ with } M_1 = \langle v_1 \rangle \oplus \langle v_2 \rangle, M_2 = \langle v_3 \rangle \oplus \langle v_4 \rangle, M_3 = \langle v_5 \rangle \oplus \langle v_6 \rangle;$$

$$P = M_1 \oplus M_2 \text{ with } M_1 = \langle v_1, v_3 \rangle \oplus \langle v_2, v_4 \rangle, M_2 = \langle v_5 \rangle \oplus \langle v_6 \rangle;$$

$$P = M_1 \text{ with } M_1 = \langle v_1, v_3, v_5 \rangle \oplus \langle v_2, v_4, v_6 \rangle.$$

In a similar way, one can show that the only time we will get uniqueness (in the same sense as for the Weyr form) for general Jordan decompositions of a nilpotent linear transformation $t : P \rightarrow P$ is when the true Jordan structure of t has no repeated basic blocks.

For another slant on what is different about the three decompositions of our particular $t : F^6 \rightarrow F^6$, let's examine them in terms of matrix representations. In the same way one gets the Jordan matrix form from a Jordan vector space decomposition into 1-dimensional subspaces (as discussed earlier), each of the three decompositions suggests a matrix representation of t . The first decomposition gives the correct Jordan matrix form J (above) for t relative to the basis $\{v_1, v_2, v_3, v_4, v_5, v_6\}$. From the second decomposition and its

suggested (ordered) basis $\{\nu_1, \nu_3, \nu_2, \nu_4, \nu_5, \nu_6\}$, we get a sort of “Jordan–Weyr” hybrid matrix

$$H = \left[\begin{array}{cc|cc|cc} 0 & 0 & 1 & 0 & & \\ 0 & 0 & 0 & 1 & & \\ \hline 0 & 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & & \\ \hline & & & & 0 & 1 \\ & & & & 0 & 0 \end{array} \right].$$

The third decomposition suggests the basis $\{\nu_1, \nu_3, \nu_5, \nu_2, \nu_4, \nu_6\}$ and this gives the correct Weyr matrix form for t

$$W = \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 1 & 0 & 0 \\ & 0 & 0 & 0 & 1 & 0 \\ & & 0 & 0 & 0 & 1 \\ \hline & & & 0 & 0 & 0 \\ & & & & 0 & 0 \\ & & & & & 0 \end{array} \right].$$

The three matrix representations J, H , and W are fundamentally different (although, of course, similar via permutation conjugations).

Returning to our proposed general formulation in Definition 4.9.1 of a Jordan decomposition of a nilpotent endomorphism $t : P \rightarrow P$, suppose we were to also insist that $m_1 > m_2 > \cdots > m_s$. Such decompositions will exist if the powers of t are regular, as we demonstrated earlier using the triangular scheme of B_{ij} . What is interesting is that now we do also get uniqueness in the strong sense of such decompositions for a nilpotent linear transformation $t : P \rightarrow P$. (By this we mean that if $P = M'_1 \oplus M'_2 \oplus \cdots \oplus M'_u$, with $M'_i = N'_{i1} \oplus N'_{i2} \oplus \cdots \oplus N'_{ip_i}$ for $i = 1, \dots, u$, is a second Jordan decomposition for t and with $p_1 > p_2 > \cdots > p_u$, then $s = u$, $m_i = p_i$ for $i = 1, \dots, s$ and $\dim N_{ij} = \dim N'_{ij}$ for all i, j .) One can deduce this uniqueness from the uniqueness of the Weyr form of t , although we won't go through the details.³¹ (Essentially one constructs a Weyr form from the associated N_{ij} as we did in the proof of Theorem 4.8.2.) However, even these sorts of decompositions will rarely agree with the true Jordan decompositions of t (except when the basic Jordan blocks are of different sizes). So no

31. This is a very good exercise. Two of the authors flunked it at their first attempt.

joy there in general. Although probably of marginal interest,³² one obtains yet another canonical form for matrices via this particular type of Jordan decomposition for nilpotent transformations t . Namely, in the nilpotent case, one first takes the Jordan matrix form $\text{diag}(J_1, J_2, \dots, J_s)$ for t , where the J_i are the basic Jordan blocks (in decreasing order of size). Then one replaces each batch of repeated diagonal blocks by the Weyr form of the corresponding submatrix. For instance, if t has Jordan structure $(5, 4, 4, 2, 2, 2)$ and a Jordan matrix $\text{diag}(J_1, J_2, J_3, J_4, J_5, J_6)$, the new canonical form is $\text{diag}(J_1, W_1, W_2)$ where W_1 and W_2 are in Weyr form with Weyr structures $(2, 2, 2, 2)$ and $(3, 3)$, respectively. What characterizes the new canonical matrices is that they are block diagonal, $\text{diag}(K_1, K_2, \dots, K_u)$, where the K_i are nilpotent homogeneous Weyr matrices whose nilpotent indices are strictly decreasing. However, why one would be interested in such a form is not clear to the authors.

Another way of restoring uniqueness to Jordan decompositions in Definition 4.9.1 (when restricted to linear transformations) would be to insist that the summands N_{ij} be indecomposable or cyclic (or both). That, however, greatly restricts the class of applicable modules P .

These attempts at being even-handed with the Jordan form have gone on long enough. It's judgment time. Despite our best efforts, we have been unable to formulate a Jordan form in anywhere near the same generality that we have achieved for the Weyr form. Accordingly, we declare the Weyr form an outright winner over the Jordan form in the module department.

In summary, even though the Jordan and Weyr forms can be derived from each other for matrices over an algebraically closed field, the above discussion suggests that the Weyr form lives in a somewhat bigger universe. It also suggests that the concept of the Weyr form of a matrix over a field is a little more "basis-free" than its Jordan counterpart, that is, its description does not need to reference a basis or 1-dimensional subspaces. This may partly explain why the Weyr form is an easier tool to use in some applications, such as those we study in later chapters.

4.10 NILPOTENT ELEMENTS WITH REGULAR POWERS

This section is only indirectly connected to the Jordan and Weyr forms, but it does provide a good ring-theoretic insight into the nature of a key condition we used in Section 4.8, namely, that *all* the powers of a nilpotent ring element are regular. For the reader, it is an optional section within an optional chapter. Accordingly, the pace and sophistication pick up appreciably.



32. One could mistakenly have made a similar comment about the Weyr form in 1885.

We require good facility in working with direct sum decompositions and module homomorphisms of a ring R regarded as a left R -module. We also need the matrix representation of an endomorphism of a free module over a commutative ring relative to a chosen basis. (See Remark 4.3.2.) So checking all the details may represent a bit of a challenge to some readers. (But isn't that what life is about?)

We aim to show, in Theorem 4.10.2 below, that in an arbitrary algebra A over a commutative ring Λ , if all powers of the nilpotent element a are regular, then at least “locally,” a looks like “a matrix in Jordan form” (or Weyr form if one wishes). The theorem was first developed in 2004 by Beidar, O'Meara, and Raphael (whilst working on a problem not related to canonical forms). Note we are not insisting on any finiteness conditions on A whatsoever (such as finite-dimensionality or chain conditions on left ideals). Our description depends on the particular ring Λ over which we choose to regard A as a Λ -algebra. Ideally, of course, we would like Λ to be a field F because that yields the best description of the element a . That may not be possible. A fallback that is always available for any ring A is to regard A as an algebra over the ring \mathbb{Z} of integers. Even here our description of a is still revealing. The reader should pay particular attention to the fact that we require *all* the powers of a to be regular, not just a itself. In the next section, we give an example to show our conclusions can't possibly hold without this full assumption.

We begin by spelling out a little result that was alluded to in Remark 4.4.7.

Lemma 4.10.1

Let R be a ring (with identity) and let $M = R$ as a left R -module. Then each R -module homomorphism $f : M \rightarrow M$ is given by right multiplication by a unique element $a \in R$, that is,

$$f(m) = ma \text{ for all } m \in M.$$

Moreover, $\text{im}(f) = Ra$, the principal left ideal of R generated by a , and $\ker(f) = \text{ann}_R(a) = \{x \in R : xa = 0\}$, the left annihilator ideal in R of a .

Proof

Let $a = f(1)$. Then for all $m \in M$

$$f(m) = f(m1) = mf(1) = ma,$$

establishing the existence of a . Uniqueness is evident upon setting $m = 1$. The other statements are clear. (The result fails for rings without identity—just consider the identity mapping.) \square

Theorem 4.10.2 (Beidar, O'Meara, and Raphael)

Let A be an algebra over a commutative ring Λ , and let $a \in A$ be a nilpotent element of index r . Assume that all the powers of a are regular in the ring A . Then there exist:

- (1) a generalized inverse b of a ,
- (2) ideals I_1, I_2, \dots, I_s of the ring Λ ,
- (3) an algebra isomorphism

$$\theta : \Lambda[a, b] \rightarrow \prod_{i=1}^s M_{m_i}(\Lambda/I_i)$$

of the Λ -subalgebra $\Lambda[a, b]$ of A generated by a and b onto the direct product of various full matrix algebras $M_{m_i}(\Lambda/I_i)$ over the factor rings Λ/I_i ,

such that $r = m_1 > m_2 > \dots > m_s \geq 1$ and $\theta(a) = (J_1, J_2, \dots, J_s)$ where

$$J_i = \begin{bmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix}.$$

The image of b has the transpose of these matrices as its components. Conversely, these conditions imply that all powers of a are regular.

Proof

We pick up on some of the ideas in the proof of Theorem 4.8.2. In the notation used there, let $R = A$, $P = R$, and let $t : P \rightarrow P$ be the map given by right multiplication by a (that is, $t(x) = xa$ for all $x \in P$). Now P is a projective left R -module by Proposition 4.3.4, and t is a nilpotent R -endomorphism of P of index r . Also the powers t^k are regular in $\text{End}_R(P)$ because if $z \in A$ is a quasi-inverse of a^k (that is, $a^k = a^k z a^k$), then right multiplication by z will supply a quasi-inverse of t^k . Therefore, as in the proof of Theorem 4.8.2, we can produce a direct sum decomposition

$$A = P = \bigoplus_{1 \leq i \leq j \leq r} B_{ij} \quad (5)$$

in which the summands are connected as in the scheme

$$\begin{array}{ccccccc}
 B_{11} & \leftarrow & B_{12} & \leftarrow & B_{13} & \leftarrow & \cdots & \leftarrow & B_{1r} \\
 & & \oplus & & \oplus & & & & \oplus \\
 & & B_{22} & \leftarrow & B_{23} & \leftarrow & \cdots & \leftarrow & B_{2r} \\
 & & & & \oplus & & & & \oplus \\
 & & & & B_{33} & \leftarrow & \cdots & \leftarrow & B_{3r} \\
 & & & & & & & & \oplus \\
 & & & & & & & & \vdots \\
 & & & & & & & & \oplus \\
 & & & & & & & & B_{rr}
 \end{array}$$

In this scheme, the arrows represent R -isomorphisms under the restriction of t , and the kernel of t is the sum $B_{11} \oplus B_{22} \oplus \cdots \oplus B_{rr}$ of the diagonal summands. The use of the notation P , R , and t is only temporary in order to make this earlier connection with nilpotent endomorphisms of a projective module. We now return to the algebra A itself.

Note that the summands B_{ij} are left ideals of A , and by Lemma 4.10.1 the left annihilator of the element a in the ring A is the kernel of t and so given as

$$\text{ann}_A(a) = B_{11} \oplus B_{22} \oplus \cdots \oplus B_{rr}.$$

With reference to the above scheme, let $b \in A$ be the element whose right multiplication map of A is the A -homomorphism that reverses the arrows (provides their inverses) and whose kernel ($= \text{ann}_A(b)$) is the last column $B_{1r} \oplus B_{2r} \oplus \cdots \oplus B_{rr}$. Here we are making use of Proposition 4.2.6 and Lemma 4.10.1. Clearly, $a = aba$ and $b = bab$ because the right multiplication maps by each side of the respective equations agree on all the summands B_{ij} . Thus, b is a generalized inverse of a .

Recall from Proposition 4.2.4 that a left ideal that is a direct summand of a ring must be a principal left ideal (in fact generated by an idempotent, although we don't need that here). Hence, we can choose $c_1, c_2, \dots, c_r \in A$ such that $B_{jr} = Ac_j$ for $j = 1, 2, \dots, r$ (i.e., we pick generators for the last column summands). Some of the B_{kr} could be 0, equivalently $c_k = 0$. We remark that $B_{1r} \neq 0$, otherwise $a^{r-1} = 0$, which would contradict a having nilpotent index r . Suppose B_{kr} is among the nonzero. Let V_k be the Λ -submodule of A generated by

$$c_k, c_k a, c_k a^2, \dots, c_k a^{r-k-1}, c_k a^{r-k}.$$

Notice the action of a under right multiplication on these generators (in the order they are presented). It forward shifts each generator to its successor and annihilates the last. Similarly, b under right multiplication annihilates the first generator and backward shifts the others to their predecessors. This is most easily seen by glancing

back at the above scheme, remembering the arrows for a and implicit for b represent inverse maps and that their left annihilators are, respectively, the main diagonal and the last column. For instance, $(c_k a^2)b = c_k a$ because a maps $c_k a \in B_{k,r-1}$ to $c_k a^2 \in B_{k,r-2}$ and so b must undo this. It follows that V_k is invariant under right multiplication by both a and b , and therefore invariant under right multiplication by $\Lambda[a, b]$. This allows us to view V_k as a right $\Lambda[a, b]$ -module, with the module action coming from right multiplication.

From the independence of the B_{ij} , we see that as a left Λ -module

$$V_k = \Lambda c_k \oplus \Lambda c_k a \oplus \Lambda c_k a^2 \oplus \cdots \oplus \Lambda c_k a^{r-k}. \quad (6)$$

Let $L_k = \text{ann}_\Lambda(c_k) = \{\lambda \in \Lambda : \lambda c_k = 0\}$ be the left annihilator of c_k in Λ . Note that, since Λ is commutative, L_k is a two-sided ideal of Λ . Furthermore, L_k is also the left annihilator in Λ of $c_k a^i$ for each $i = 0, 1, \dots, r-k$, since $c_k a^i b^i = c_k$ and A is an algebra over Λ . We can therefore conclude that the decomposition in (6) displays V_k as a free Λ/L_k module with a basis $\mathcal{B}_k = \{c_k, c_k a, c_k a^2, \dots, c_k a^{r-k}\}$. (Here the module action of $\lambda + L_k \in \Lambda/L_k$ on an element $v \in V_k$ is $(\lambda + L_k) \cdot v = \lambda v$.) The number of elements in this basis is $n_k = r - k + 1$.

Let

$$\eta_k : \Lambda[a, b] \rightarrow M_{n_k}(\Lambda/L_k)$$

be the Λ -algebra homomorphism that is the composition of first representing an element of $\Lambda[a, b]$ as a Λ/L_k endomorphism of V_k via right multiplication, and then representing that endomorphism as an $n_k \times n_k$ matrix over Λ/L_k relative to the basis \mathcal{B}_k . (See Remark 4.3.2. However, because of the “anti-isomorphic twist” that comes from right multiplying, we need to choose the transposes of the representing matrices.) From the shift effects of a and b on the basis elements, we have

$$\eta_k(a) = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 0 & 1 \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix} \quad (7)$$

and $\eta_k(b)$ is the transpose of this matrix.

The construction of η_k for a fixed k was premised on B_{kr} being nonzero. Let's now consider all such k , and index them, say, as $k_1 = 1 < k_2 < \cdots < k_s$. It pays to

relabel a few earlier items: set $\mathcal{C}_i = \mathcal{B}_{k_i}$, $I_i = L_{k_i}$, $m_i = n_{k_i}$, and $\theta_i = \eta_{k_i}$. We then have Λ -algebra homomorphisms

$$\theta_i : \Lambda[a, b] \rightarrow M_{m_i}(\Lambda/I_i)$$

for $i = 1, 2, \dots, s$ such that $\ker(\theta_i) = \text{ann}_{\Lambda[a, b]}(\mathcal{C}_i)$ and with the matrices $\theta_i(a)$ and $\theta_i(b)$ as displayed (or described) in the previous paragraph. We can tie these together to produce a Λ -algebra homomorphism

$$\theta : \Lambda[a, b] \rightarrow \prod_{i=1}^s M_{m_i}(\Lambda/I_i)$$

by letting $\theta(x) = (\theta_1(x), \theta_2(x), \dots, \theta_s(x))$ for all $x \in \Lambda[a, b]$. Its kernel is the right annihilator of all the \mathcal{C}_i . But our scheme of B_{ij} shows that $\mathcal{C}_1 \cup \mathcal{C}_2 \cup \dots \cup \mathcal{C}_s$ generates A as a left ideal (because that left ideal contains all the B_{ij}). Thus, $\ker(\theta) = 0$ and so θ maps $\Lambda[a, b]$ isomorphically into the product of full matrix algebras. To complete our proof, we need to show that θ is an onto map.

It is a simple (and cute) exercise to show that over a commutative ring R (with identity), $M_n(R)$ can be generated as an R -algebra by the matrix

$$A = \begin{bmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix}$$

and its transpose B . The trick is to first produce some matrix unit in terms of A and B (such as $e_{11} = I - BA$) and then use the shifting effects of A and B under repeated left and right multiplications to round up all the other matrix units e_{ij} .³³ Simply observe that **left** multiplying a matrix unit by A (resp. B) moves it up (resp. down) one place, while **right** multiplying the matrix unit by A (resp. B) moves it right (resp. left) one place. We will make use of this idea.³⁴ Remember that the i th component of $\theta(a)$ (that is $\theta_i(a)$), as an $m_i \times m_i$ matrix, has the form in (7), and the i th component for $\theta(b)$ is its transpose. Also $r = m_1 > m_2 > \dots > m_s \geq 1$. Hence, $\theta(a)^{r-1} = (e_{1r}, 0, 0, \dots, 0)$ where e_{1r} is the matrix unit in $M_r(\Lambda/I_1)$ with a 1 in the $(1, r)$ position and zeros elsewhere. We can now use repeated multiplications by $\theta(a)$ and $\theta(b)$ to get all the matrix

33. We could show directly that $e_{ij} = B^{j-1}A^{i-1} - BA^i$ but we need a more subtle approach.

34. Frankie Laine sums up the idea in the lyrics "Move 'em out, head 'em up, Rawhide."

units $(e_{ij}, 0, 0, \dots, 0)$ in the image of θ . Therefore all elements of the form $(x, 0, 0, \dots, 0)$, where $x \in M_{m_1}(\Lambda/I_1)$, are in the image of θ . In particular, the first components of $\theta(a)$ and $\theta(b)$ are in the image. Subtract these from $\theta(a)$ and $\theta(b)$. Now repeat the same argument with the second components of this new pair of elements in $\prod_{i=1}^s M_{m_i}(\Lambda/I_i)$, and so on. In this way, we can show the image of θ contains the full direct product of the matrix algebras $M_{m_i}(\Lambda/I_i)$. (One point to watch, however, is when $m_s = 1$, and for that use the fact that $\theta(\Lambda[a, b])$ must contain the identity.) Thus, we have established the necessity of (1), (2), and (3).

To establish the converse of our theorem, suppose there is an isomorphism

$$\theta : \Lambda[a, b] \rightarrow \prod_{i=1}^s M_{m_i}(\Lambda/I_i)$$

as described for a suitable generalized inverse b of a . In the image, one can quickly see that $\theta(b)^i$ is a quasi-inverse of $\theta(a)^i$ for $i = 1, 2, \dots, r-1$. (Just check this in each component by the standard shifting arguments.) Therefore, since θ is an isomorphism, b^i is a quasi-inverse of a^i for $i = 1, 2, \dots, r-1$. Hence, all the powers of a must be regular. \square

Notice that if the ring Λ in Theorem 4.10.2 is a field F (the best case), then we can replace the factor rings Λ/I_i by F , because a field has no proper nonzero ideals. The conclusion in 4.10.2 then really does put $\theta(a)$ in Jordan form, once we identify $(A_1, A_2, \dots, A_s) \in \prod_{i=1}^s M_{m_i}(F)$ with the block diagonal matrix $\text{diag}(A_1, A_2, \dots, A_s)$. This gives $\theta(a)$ the Jordan structure (m_1, m_2, \dots, m_s) with $m_1 > m_2 > \dots > m_s$. It is a little curious that the Jordan blocks are all distinct, particularly since a special case of the theorem would be to start with the algebra $A = M_n(F)$ and a nilpotent matrix $a \in A$ in Jordan form but with some repeated Jordan blocks. There is nothing contradictory about this, however. We leave it to the reader to resolve this apparent contradiction to the uniqueness of the Jordan form.

Finally, we close this section with a couple of remarks, hopefully thought-provoking. The curious reader may also wish to check out some of the longstanding, simply stated, open problems in Goodearl's *Von Neumann Regular Rings* concerning unit-regularity and direct finiteness. And by all means, feel free to solve some of them!³⁵

35. For instance, Open Problem 3 asks if every element of a directly finite, simple regular ring must be unit-regular. (G.M. Bergman showed by a clever argument in the 1970s that the answer is "no" if one drops simplicity. See Example 5.10 of Goodearl's text.) A negative answer to Problem 3 would also give a negative answer to the fundamental Separativity Problem mentioned in footnote 17 on p. 159.

Remarks 4.10.3

- (1) Even in the case of a regular algebra A over a field F , a general element $a \in A$ that is not nilpotent certainly need not sit inside a subalgebra B of A that is isomorphic to some finite direct product of matrix algebras $M_{m_i}(F)$. For this would have ring-theoretic consequences: a would have to be a unit-regular element of B (possess an invertible quasi-inverse), hence a unit-regular element of A . And the algebra A would be “directly finite,” in the sense that any element with a one-sided inverse would have a two-sided inverse. Thus, the algebra A of all linear transformations of an infinite-dimensional vector space supplies an easy counterexample (see Corollary 4.4.8). But can something still be salvaged outside the nilpotent case?
- (2) By Theorem 4.10.2 and an extension of the argument in (1), if a nilpotent element a of a completely general algebra A (over any commutative ring) has all its powers regular, then a is unit-regular. In particular, nilpotent elements of a regular algebra A are unit-regular. But regular algebras can contain elements that are not unit-regular. So what has happened to our fundamental “reduction to the nilpotent case” principle, which we use repeatedly for (finite) matrices throughout our book? Again, can something still be salvaged, say for certain non-nilpotent elements of an infinite-dimensional regular algebra over a field?

4.11 A REGULAR NILPOTENT ELEMENT WITH A BAD POWER

We present an example of a nilpotent element a in a finite-dimensional algebra A such that a is regular but its square is not. The example appeared in a 1995 paper by Yu, although Yu credits it to a communication by Goodearl.

Example 4.11.1

Let F be a field and let $R = F[x]/(x^2)$ be the ring of polynomials over F modulo x^2 . Let $A = M_2(R)$ be the algebra of 2×2 matrices over R . Then the matrix

$$a = \begin{bmatrix} 0 & 1 \\ 0 & x \end{bmatrix}$$

is a regular nilpotent element of A but its square is not regular.

We denote the entries of matrices in A by polynomials but the calculations in matrix operations are to be done modulo x^2 (setting x^2 and higher powers to 0). Thus,

$$a^2 = \begin{bmatrix} 0 & x \\ 0 & x^2 \end{bmatrix} = \begin{bmatrix} 0 & x \\ 0 & 0 \end{bmatrix}$$

and

$$a^3 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

so a is nilpotent of index 3. One checks directly that

$$u = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

is a quasi-inverse of a ($a = aua$), whence a is a regular element of A . (In fact, since u is a unit, a is unit-regular, an even nicer property.) Suppose

$$c = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$$

is a quasi-inverse of a^2 (that is, $a^2 = a^2ca^2$). Then

$$\begin{aligned} \begin{bmatrix} 0 & x \\ 0 & 0 \end{bmatrix} &= \begin{bmatrix} 0 & x \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p & q \\ r & s \end{bmatrix} \begin{bmatrix} 0 & x \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} xr & xs \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & x \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & x^2r \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned}$$

which is a contradiction. Therefore, a^2 is not regular in A .

For the seasoned ring theorist,³⁶ there is a slicker argument showing a^2 can't be regular. One observes that a^2 is in the Jacobson radical $J(A)$ of A (here coinciding with the maximum nilpotent ideal $M_2((x)/(x^2))$). If a^2 were regular with quasi-inverse c , then $e = a^2c$ would be a nonzero idempotent in $J(A)$, impossible because $1 - e$ is not invertible.

So we can't get by in Theorem 4.10.2 by assuming only a is regular, even for an algebra over a field. It is a tight theorem. \square

We have now completed the theory side of the Weyr form. We are primed and ready for applications.

36. Or those readers who are still with us !

BIOGRAPHICAL NOTE ON VON NEUMANN

The mathematician Jean Dieudonné once described von Neumann as the “last of the great mathematicians.” John von Neumann was born János von Neumann in Budapest, Hungary, on December 28, 1903, the son of a successful banker. Although the family was Jewish, they were not strict observers and John entered the Lutheran Gymnasium in 1911. His prodigious memory and mathematical ability were apparent at an early age. At six, he could divide two eight-digit numbers in his head; by eight, he had mastered calculus; by twelve, he had reached the graduate level in mathematics. He completed his Gymnasium studies in 1921, but his father wanted him to pursue a business career instead of mathematics. A compromise was reached: chemistry. However, Hungarian university limitations on Jews saw John enrolling at the University of Berlin in 1921, before transferring to the Technische Hochschule Zurich in 1923 to study mathematics. He went on to receive his doctorate in mathematics from the University of Budapest in 1926, following this with a postdoctoral Rockefeller Scholarship at the University of Göttingen where he studied under Hilbert. With a rapidly rising academic status, he became a visiting lecturer at Princeton University in 1930 and a professor there in the following year. His early research involved mathematical logic, axiomatic set theory, computer science, and measure theory. He is one of the great pioneers of modern computer science (some even describe him as the father of the modern computer), being the first to formulate the concept of a stored computer program (thereby enabling a computer to perform different tasks without “rewiring”). His classical 1932 text *Mathematische Grundlagen der Quantummechanik* set the foundations for quantum theory and statistical mechanics. With F. J. Murray, he wrote a series of papers in the late 1930s on what are now called von Neumann algebras. His ability to develop ground-breaking results in a variety of fields continued with, for example, the 1944 text *Theory of Games and Economic Behavior*. Von Neumann’s scientific achievements, however, have never really received widespread popular recognition (compared, say, with his contemporary Einstein), possibly because he was one of the principal players in the Manhattan project of World War II and in the postwar development of the hydrogen bomb. During his life he received many honors, including two (U.S.) Presidential Awards. He died on February 8, 1957, in Washington, D.C. It is a commentary on the times that while heavily medicated and dying from cancer, he was placed under military security lest he reveal military secrets.

This page intentionally left blank

PART TWO

Applications of the Weyr Form

The applications we have chosen have a common thread—they all involve commuting matrices over an algebraically closed field. But the basic questions studied within the next three chapters are essentially quite different, as are the techniques used to answer them. The chapters are largely independent and can be read as such.

This page intentionally left blank

Gerstenhaber's Theorem

The time has come to put the Weyr form to work. Our first application is perhaps more on the “pure” side of linear algebra. It concerns some special cases of a rather difficult but interesting problem: bounding the dimension of a commutative subalgebra \mathcal{A} of $M_n(F)$. We discussed aspects of this in Section 3.5 of Chapter 3, including Schur’s sharp upper bound of $\lfloor n^2/4 \rfloor + 1$ for a general commutative subalgebra. However, this bound does not take into account the number of generators required for \mathcal{A} . We will see that, in some situations, knowing the number of generators can lead to a considerable improvement on Schur’s bound.

Fix a field F and consider the algebra $M_n(F)$ of all $n \times n$ matrices over F . It is a simple consequence of the Cayley–Hamilton theorem that the subalgebra $F[A]$ generated by a single matrix A can have dimension at most n . If we allow two matrices A and B , then with the right choice, the subalgebra $F[A, B]$ can be all of $M_n(F)$ and so has dimension n^2 . (For example, we can take A to be a basic nilpotent Jordan matrix and B its transpose; see proof of Theorem 4.10.2.) Thus, it is most surprising that if we require A and B to *commute*, then as with the one generator case, the dimension of $F[A, B]$ still cannot exceed n . This is the content of a 1961 theorem of Gerstenhaber. The theorem has been re-derived by a number of authors over the years, some using powerful techniques of algebraic geometry (as did Gerstenhaber himself),

others using purely matrix-theoretic arguments involving the Jordan form. In this chapter, we offer a proof that utilizes both the Jordan and Weyr forms, resulting in a short, transparent proof of Gerstenhaber's theorem, with the bonus of an explicit spanning set for $F[A, B]$ in terms of the Weyr structure of A when A is nilpotent (the core case). Our spanning set result is the "dual" of a description given by Barriá and Halmos, and Laffey and Lazarus, in terms of the Jordan structure of A .

Our proof of Gerstenhaber's theorem is along the lines of the Barriá and Halmos proof in 1990. There are three steps, which we cover in Sections 5.1, 5.2, and 5.3: (1) reduction to the case where A and B are commuting nilpotent $n \times n$ matrices, (2) a generalized Cayley–Hamilton equation, and (3) an inductive step to smaller-size matrices. Since the linear independence of matrices over F is retained when passing to a larger field, there is no loss of generality in establishing Gerstenhaber's theorem by assuming F is algebraically closed. Henceforth, we do this.

In Section 5.4, we study the 2-generated maximal commutative subalgebras of $M_n(F)$, which were independently characterized in the early 1990s by Laffey and Lazarus, and Neubauer and Saltman. We show how the Weyr form leads to a relatively simple proof of their result in the homogeneous case.

In the case of commutative 3-generated subalgebras of $M_n(F)$, it is an open question as to whether n is still the best upper bound for dimension. In Section 5.5 we show how the Weyr form suggests new techniques for tackling the 3-generator case, in terms of the concept of a "pullback" of a matrix that centralizes a given nilpotent Weyr matrix. The leading edge subspaces introduced in Section 3.4 of Chapter 3 play a critical role in the arguments of the current Sections 5.4 and 5.5. The same arguments formulated in terms of the Jordan form would be unnatural and unwieldy.

At one point in Chapter 6, as an offshoot of a seemingly unrelated and more "applied" development, we discuss some sharp bounds on the dimension of a commutative subalgebra \mathcal{A} of $M_n(\mathbb{C})$ in terms of n and the " d -regularity" of one of its members. These bounds, for example, $5n/4$ when $d = 2$, are much smaller than Schur's but are still independent of the number of generators.

The Weyr form should not be regarded as being in "competition" with the Jordan form. In some situations (usually involving a single matrix or transformation in isolation) the Jordan form is more useful, whilst in others (typically involving the interaction of several matrices) the Weyr form is better. This chapter illustrates the advantage of being prepared to switch back and forth between the two forms, utilizing the duality in Section 2.4 of Chapter 2. Some might argue that since the two forms are conjugate under a known permutation transformation, this amounts to just using "smoke and mirrors." Our arguments suggest not—a known proof for one form does not always transform to a direct

proof in terms of the other. And even when both forms do the job, one form can be more natural and suggestive than the other.

5.1 k -GENERATED SUBALGEBRAS AND NILPOTENT REDUCTION

Studying subalgebras \mathcal{A} of $M_n(F)$ doesn't sound too taxing, until one realizes that this study encompasses *all* finite-dimensional algebras \mathcal{A} over F .¹ Indeed, an arbitrary algebra \mathcal{A} of dimension m can be isomorphically embedded in $M_m(F)$ through its “regular representation.” Specifically, let V be \mathcal{A} as a vector space, and choose a basis \mathcal{B} for V . We have an algebra embedding

$$\begin{aligned}\theta : \mathcal{A} &\longrightarrow \mathcal{L}(V) \\ \theta(X)(Y) &= XY \text{ for all } X \in \mathcal{A} \text{ and } Y \in V\end{aligned}$$

of \mathcal{A} into the algebra $\mathcal{L}(V)$ of linear transformations of V , which we can follow up with the algebra isomorphism $\mathcal{L}(V) \rightarrow M_m(F)$, $T \mapsto [T]_{\mathcal{B}}$ to yield an algebra embedding ψ of \mathcal{A} into $M_m(F)$. For instance, if $F = \mathbb{R}$ and $\mathcal{A} = \mathbb{R}[i, j, k]$ is the algebra of real quaternions (where $i^2 = j^2 = k^2 = -1$, $ij = k$, $jk = i$, $ki = j$, $ji = -k$, $kj = -i$, $ik = -j$), and we take $\mathcal{B} = \{1, i, j, k\}$, we obtain the regular representation

$$\begin{aligned}\psi : \mathcal{A} &\longrightarrow M_4(\mathbb{R}) \\ a_0 + a_1i + a_2j + a_3k &\longmapsto \begin{bmatrix} a_0 & -a_1 & -a_2 & -a_3 \\ a_1 & a_0 & -a_3 & a_2 \\ a_2 & a_3 & a_0 & -a_1 \\ a_3 & -a_2 & a_1 & a_0 \end{bmatrix}.\end{aligned}$$

Thus, the image of ψ is a subalgebra of $M_4(\mathbb{R})$ which is an isomorphic copy of the real quaternions.² A similar warning is in place when we study, say, 2-generated commutative subalgebras of a general $M_n(F)$ —they encompass *all* 2-generated commutative algebras of finite dimension.

Let \mathcal{C} be an algebra (with identity) over our field F . For the moment, \mathcal{C} need not be finite-dimensional or even commutative. If A_1, A_2, \dots, A_k are k fixed

1. This is entirely analogous to studying subgroups of a general symmetric group S_n . By Cayley's theorem, they encompass all finite groups.

2. Here is yet another illustration of a matrix outcome achieved effortlessly through the use of linear transformations. This supports the authors' “philosophical” comments made in the introduction to Chapter 1.

members of \mathcal{C} , as before we denote by $F[A_1, A_2, \dots, A_k]$ the subalgebra of \mathcal{C} generated by A_1, A_2, \dots, A_k , that is, the smallest subalgebra of \mathcal{C} containing A_1, A_2, \dots, A_k . We refer to $F[A_1, A_2, \dots, A_k]$ as a k -generated subalgebra. For general algebras, finitely generated subalgebras can be very complicated.³ About the only thing we can say in general about $F[A_1, A_2, \dots, A_k]$ is that it is spanned as a vector space by the (noncommutative) words in A_1, A_2, \dots, A_k . Thus, for $k = 2$, a typical member of $F[A_1, A_2]$ might be

$$7 + 5A_2 + 2A_1A_2 - A_2A_1 + 4A_2^3A_1^4A_2 - 6A_1A_2A_1A_2^2A_1 + A_1^3A_2^2A_1^2.$$

However, if A_1, A_2, \dots, A_k commute, then the description of the members of $F[A_1, A_2, \dots, A_k]$ is much simpler. They are polynomials in the generators. The typical member above in the commutative case then becomes

$$7 + 5A_2 + A_1A_2 + 4A_1^4A_2^4 - 6A_1^3A_2^3 + A_1^5A_2^2.$$

Also, if the algebra \mathcal{C} is finite-dimensional, then each of its subalgebras is k -generated for some k (because one can take a vector space basis of the subalgebra as a set of generators), but, of course, k is not unique. For the most tractable description, one aims for the smallest k . If $k = 1$, then we have a complete description of $F[A_1]$ as soon as we know the minimal polynomial of A_1 . Nothing complicated there. So the first interesting case is when $k = 2$.

For the remainder of the chapter, we restrict our parent algebra to $\mathcal{C} = M_n(F)$, the algebra of $n \times n$ matrices over F . We will be interested in *commutative* k -generated subalgebras \mathcal{A} of \mathcal{C} , particularly for $k = 2$ and $k = 3$. With the former case, much is known (such as Gerstenhaber's theorem, to come), but with the latter, there are still open problems. One philosophical point: in studying k -generated subalgebras, are we interested in describing a given subalgebra in terms of suitable generators, or is there an intrinsic interest in studying some k given matrices by looking at the subalgebra they generate? The answer can be either, depending on the circumstances. For instance, in Chapter 6, when we investigate an approximate simultaneous diagonalization problem, our primary interest lies in the k matrices we start with, and we can use properties of the subalgebra they generate (such as its dimension) to shed light on the approximation problem.

Next we aim to establish the expected result, that without loss of generality, we can assume the commuting matrices A_1, A_2, \dots, A_k are all nilpotent.

3. Any countable-dimensional commutative algebra can be made the center of a suitable 2-generated algebra. (See the 1989 paper by O'Meara, Vinsonhaler, and Wickless.) Consequently, there are 2^{\aleph_0} nonisomorphic 2-generated algebras over even a countable field, compared with only \aleph_0 nonisomorphic 1-generated algebras.

Proposition 5.1.1

Suppose A_1, A_2, \dots, A_k are commuting $n \times n$ matrices over an algebraically closed field F . Then there is a simultaneous similarity transformation of A_1, A_2, \dots, A_k such that:

- (1) All the A_i become block diagonal with matching block structures and such that each diagonal block has only a single eigenvalue (ignoring multiplicities). That is, relative to some diagonal block sizes m_1, m_2, \dots, m_t ,

$$A_i = \text{diag}(A_{i1}, A_{i2}, \dots, A_{it}) \text{ for } i = 1, 2, \dots, k$$

where A_{ij} is an $m_j \times m_j$ matrix with a single eigenvalue.

- (2) As algebras,

$$F[A_1, A_2, \dots, A_k] \cong \prod_{j=1}^t F[A_{1j}, A_{2j}, \dots, A_{kj}].$$

- (3) For a fixed j , after the subtraction of suitable scalar matrices, the k commuting $m_j \times m_j$ matrices $A_{1j}, A_{2j}, \dots, A_{kj}$ become nilpotent (but generate the same subalgebra).

Proof

This is a standard argument involving generalized eigenspaces. Let $\lambda_1, \lambda_2, \dots, \lambda_t$ be the distinct eigenvalues of A_1 and let $p(x) = (x - \lambda_1)^{m_1} (x - \lambda_2)^{m_2} \dots (x - \lambda_t)^{m_t}$ be the characteristic polynomial of A_1 . By the Corollary 1.5.4 to the generalized eigenspace decomposition, applied to A_1 , there is a similarity transformation under which $A_1 = \text{diag}(A_{11}, A_{12}, \dots, A_{1t})$, where A_{1j} is an $m_j \times m_j$ matrix having λ_j as its only eigenvalue. Perform the same similarity transformation on A_2, \dots, A_k . The new matrices A_2, \dots, A_k must still centralize the new A_1 . Therefore, by Proposition 3.1.1, A_2, A_3, \dots, A_k are also block diagonal of the same block structure as A_1 , say $A_i = \text{diag}(A_{i1}, A_{i2}, \dots, A_{it})$ for $i = 1, 2, \dots, k$.

For each $i = 1, \dots, t$ let $f_i(x) = \prod_{j \neq i} (x - \lambda_j)^{m_j}$. These polynomials are relatively prime, so for suitable polynomials $g_1(x), g_2(x), \dots, g_t(x)$ we have

$$(*) \quad f_1(x)g_1(x) + f_2(x)g_2(x) + \dots + f_t(x)g_t(x) = 1.$$

Let $E_i = f_i(A_1)g_i(A_1)$ for $i = 1, \dots, t$. Note that $E_i \in F[A_1]$. By $(*)$, we have

$$(**) \quad E_1 + E_2 + \dots + E_t = I.$$

As a block diagonal matrix, write $E_i = \text{diag}(E_{i1}, E_{i2}, \dots, E_{it})$. Since A_{1j} has λ_j as its only eigenvalue, $(A_{1j} - \lambda_j I)^{m_j} = 0$ by the Cayley-Hamilton theorem.

Thus, $f_i(A_{1j}) = 0$ for $j \neq i$. Evaluating a polynomial expression involving a block diagonal matrix can be done by evaluating the expression on each of its diagonal blocks. Therefore, $E_{ij} = f_i(A_{1j})g_i(A_{1j}) = 0$ for $j \neq i$. In conjunction with (**), this implies that E_i is the block diagonal matrix with the $m_i \times m_i$ identity matrix as its i th diagonal block and all other blocks zero. The ring import of this is that we have produced orthogonal idempotents E_1, E_2, \dots, E_t in $F[A_1]$ summing to the identity. Therefore, there is an associated algebra direct product decomposition of any commutative algebra containing A_1 , in particular, a decomposition of $\mathcal{A} = F[A_1, A_2, \dots, A_k]$. Specifically, we have an algebra isomorphism

$$\theta : \mathcal{A} \longrightarrow \prod_{j=1}^t E_j \mathcal{A}, \quad X \longmapsto (E_1 X, E_2 X, \dots, E_t X).$$

Upon identifying each $E_j X$ with its j th diagonal block (its other blocks are zero), one sees that θ induces an isomorphism

$$F[A_1, A_2, \dots, A_k] \cong \prod_{j=1}^t F[A_{1j}, A_{2j}, \dots, A_{kj}].$$

We are finished except for the point of each A_{ij} having a single eigenvalue for $i \neq 1$. But if one of these blocks A_{ij} has two distinct eigenvalues, we can repeat the splitting on that block (and induced splittings of the matching blocks of the other A s) by the same argument we have used for A_1 . Eventually (or by induction), we achieve a splitting in which each A_{ij} has only a single eigenvalue, whence it is a scalar matrix plus a nilpotent matrix. (We do not exclude the possibility that some A_{ij} and A_{im} could share a common eigenvalue when $j \neq m$.) \square

Before going on to 2-generated commutative subalgebras, let us quickly review, by way of examples, what the 1-generated subalgebras look like. (Of course, these are automatically commutative.) The reader should be warned, however, that 2-generated commutative subalgebras are more than twice as complicated.

Example 5.1.2

Let $A \in M_n(F)$. By the Cayley–Hamilton theorem, the powers A^0, A^1, \dots, A^n are linearly dependent. Let $r \leq n$ be the first positive integer such that A^r is dependent on the earlier powers, and let

$$A^r = c_0 I + c_1 A + c_2 A^2 + \dots + c_{r-1} A^{r-1}$$

be the dependence relation. Then the minimal polynomial of A is given by $m(x) = x^r - c_{r-1}x^{r-1} - \dots - c_2x^2 - c_1x - c_0$. As algebras, $F[A]$ is isomorphic to the factor algebra $F[x]/(m)$ where (m) is the principal ideal generated by $m(x)$. Moreover, $\{I, A, A^2, \dots, A^{r-1}\}$ is a basis for the 1-generated subalgebra $F[A]$. In particular, the dimension of $F[A]$ is the degree of the minimal polynomial of A . Multiplication of the basis elements is determined completely by the displayed dependence relation (together with the usual multiplication of powers $A^i A^j = A^{i+j}$ for $i + j < r$). All this is well known and easy to establish. (See Proposition 1.4.2.)

The decomposition in Proposition 5.1.1 says we can further reduce to the case where the minimal polynomial is $m(x) = x^r$ for some r (whence A has nilpotency index r). At this point, if we are interested in only the structure of $F[A]$, we can assume A is a nilpotent matrix in Jordan or Weyr form. The simplest case is when A is a basic Jordan matrix, say the 7×7

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 1 & 0 & 0 & 0 & 0 \\ & & 0 & 1 & 0 & 0 & 0 \\ & & & 0 & 1 & 0 & 0 \\ & & & & 0 & 1 & 0 \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{bmatrix}.$$

Then A has nilpotency index $r = 7$ so $\{I, A, A^2, A^3, A^4, A^5, A^6\}$ is a basis for $F[A]$. The general member of $F[A]$ therefore looks like

$$aI + bA + cA^2 + dA^3 + eA^4 + fA^5 + gA^6 = \begin{bmatrix} a & b & c & d & e & f & g \\ & a & b & c & d & e & f \\ & & a & b & c & d & e \\ & & & a & b & c & d \\ & & & & a & b & c \\ & & & & & a & b \\ & & & & & & a \end{bmatrix}.$$

Notice that with a basic Jordan matrix A , the subalgebra $F[A]$ coincides with the centralizer $\mathcal{C}(A)$ of A within the algebra $M_n(F)$ and has dimension n . See Proposition 3.2.4. Since $F[A]$ is therefore self-centralizing, any k -generated commutative subalgebra of $M_n(F)$ containing A is just the 1-generated $F[A]$. So it is not possible to build interesting examples of, say, 2-generated commutative subalgebras starting with A . Notice also that putting A in Weyr form will produce nothing new because a basic Jordan matrix is its own Weyr form with Weyr structure $(1, 1, \dots, 1)$. \square

Example 5.1.3

As a second example of a 1-generated subalgebra, suppose in Jordan form

$$A = \left[\begin{array}{ccc|cc} 0 & 1 & 0 & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & & \\ \hline & & & 0 & 1 \\ & & & 0 & 0 \\ \hline & & & & & 0 & 1 \\ & & & & & 0 & 0 \end{array} \right],$$

with Jordan structure $(3, 2, 2)$ and therefore nilpotency index $r = 3$. Then I, A, A^2 form a basis for $F[A]$, and the general member of the subalgebra looks like

$$aI + bA + cA^2 = \left[\begin{array}{ccc|cc} a & b & c & & \\ 0 & a & b & & \\ 0 & 0 & a & & \\ \hline & & & a & b \\ & & & 0 & a \\ \hline & & & & & a & b \\ & & & & & 0 & a \end{array} \right].$$

In Weyr form,

$$A = \left[\begin{array}{ccc|ccc|c} 0 & 0 & 0 & 1 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 1 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 1 & \\ \hline & & & 0 & 0 & 0 & 1 \\ & & & 0 & 0 & 0 & 0 \\ & & & 0 & 0 & 0 & 0 \\ & & & & & & 0 \end{array} \right]$$

has the dual Weyr structure $(3, 3, 1)$ and the general member of $F[A]$ looks like

$$aI + bA + cA^2 = \left[\begin{array}{ccc|ccc|c} a & 0 & 0 & b & 0 & 0 & c \\ 0 & a & 0 & 0 & b & 0 & 0 \\ 0 & 0 & a & 0 & 0 & b & 0 \\ \hline & & & a & 0 & 0 & b \\ & & & 0 & a & 0 & 0 \\ & & & 0 & 0 & a & 0 \\ & & & & & & a \end{array} \right].$$

Of the two descriptions, Jordan probably does a better job here because the matrices in $F[A]$ are block diagonal. However, the situation changes when we look at 2-generated commutative subalgebras $F[A, B]$, because then even with A in Jordan form, B is only constrained by the centralizer description in Proposition 3.1.2. So, in general, the members of $F[A, B]$ are not even block upper triangular relative to the block structure of A . It is the cleaner block upper triangular description of the centralizer of a Weyr matrix, given in Proposition 2.3.3, which seems to be the key as to why the Weyr form is more suited to the study of k -generated commutative subalgebras of matrices when $k > 1$. \square

Simple-mindedness can be a virtue in mathematics. Could it be that given two commuting matrices $A, B \in M_n(F)$, there exists a third matrix $C \in M_n(F)$ with $F[A, B] \subseteq F[C]$? Put another way, does commutativity of A and B come from both being polynomials in some common matrix C ? If so, then $\dim F[A, B] \leq \dim F[C] \leq n$ implies $\dim F[A, B] \leq n$, and we would have an easy proof of Gerstenhaber's theorem. Alas, this is not to be, as the following easy example shows.⁴

Example 5.1.4

Let

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Then A and B are commuting nilpotent matrices of nilpotency index 2 and $AB = 0$. Hence, $\{I, A, B\}$ is a basis for $F[A, B]$. Suppose $F[A, B] \subseteq F[C]$ for some $C \in M_3(F)$. Inasmuch as $\dim F[C] \leq 3$ and $\dim F[A, B] = 3$, we must have $\dim F[C] = 3$ and therefore $F[A, B] = F[C]$. As a linear combination of I, A, B , write

$$C = rI + sA + tB.$$

Then $(C - rI)^2 = (sA + tB)^2 = 0$ so the minimal polynomial of C divides $(x - r)^2$ and therefore has degree at most 2. Since the degree of the minimal polynomial equals $\dim F[C]$ (see Proposition 1.4.2), we are looking at a contradiction. Therefore A and B are not polynomials in some common matrix C . \square

4. Frobenius asked this question in 1896. This old example, in response to Frobenius, made its debut in a 1919 paper by H. B. Phillips.

5.2 THE GENERALIZED CAYLEY–HAMILTON EQUATION

Our treatment of this topic follows that of Barriá and Halmos in their 1990 paper. In turn, they attribute their version of the generalized Cayley–Hamilton equation to Ingraham and Trimble in the early 1940s, but with a simplified proof. Laffey and Lazarus also gave an independent treatment of the result in 1991. It is important to note that the generalization is not stated in terms of a specific “characteristic polynomial” equation. (It could be, but this would not look pretty.) Rather, what is of interest is the existence of a “dependence relation” involving a specific power B^d of an $n \times n$ matrix B and its lower powers, where the “coefficients” come from ordinary polynomials (over F) in some prescribed matrix A that commutes with B . For this view, one should think of the classical Cayley–Hamilton result in terms of $A = I$ and $d = n$. In the generalized version, d is usually much smaller than n . That is the whole point.

At one place in the proof, we use the Chinese remainder theorem, not in the usual integer form, but for polynomials over a field. The integer version establishes a simultaneous solution to a finite system of congruences whose moduli are relatively prime. For instance, the system

$$x \equiv 5 \pmod{6}$$

$$x \equiv 3 \pmod{10}$$

$$x \equiv 2 \pmod{21}$$

has the simultaneous solution $x = 23$. The simple proof, which is based on the fact there is an integer combination of the moduli equal to 1, can be found in most standard undergraduate algebra texts. Exactly the same proof works over any Euclidean domain D , in particular for $D = F[x]$, where F is a field. Thus, given a collection $m_1(x), m_2(x), \dots, m_t(x)$ of polynomials having no (nontrivial) common factor, a system of congruences

$$p(x) \equiv p_1(x) \pmod{m_1(x)}$$

$$p(x) \equiv p_2(x) \pmod{m_2(x)}$$

$$\vdots$$

$$p(x) \equiv p_t(x) \pmod{m_t(x)}$$

has some common solution $p(x)$, that is, $p(x) = p_i(x) + m_i(x)u_i(x)$ for some polynomials $u_i(x)$ and for $i = 1, 2, \dots, t$.

Theorem 5.2.1 (The Generalized Cayley–Hamilton Equation).⁵

Let A and B be commuting $n \times n$ matrices over F , and let d be the largest number of basic λ -Jordan blocks (i.e., basic blocks having eigenvalue λ) in the Jordan form of A , where λ ranges over the eigenvalues of A .⁶ Then

$$B^d = A_0 + A_1 B + A_2 B^2 + \cdots + A_{d-1} B^{d-1}$$

for some matrices A_0, A_1, \dots, A_{d-1} in $F[A]$ (so the A_i are polynomials in A with coefficients from F). The classical Cayley–Hamilton equation is the special case $A = I$ and $d = n$.

Proof

We break the proof into three cases: (1) A is nilpotent of homogeneous structure; (2) A is nilpotent of nonhomogeneous structure; (3) A is completely general. The first and third cases proceed smoothly (and indeed elegantly, as one comes to expect from Barría and Halmos). It is the second case where one needs to exercise care. Barría and Halmos employed the Jordan form to reduce this case to the first. We shall instead use the Weyr form but with the same objective in mind. The Weyr form makes this step slightly clearer, conceptually at least. Notice that when A is nilpotent, the integer d is just the nullity of A . Also note (by duality) that A has a homogeneous Jordan structure if and only if it has a homogeneous Weyr structure.

Case (1): A is nilpotent of homogeneous Jordan structure.

A similarity transformation of A and B won't affect the theorem, so we can assume A is a nilpotent Jordan matrix of nilpotency index r with homogeneous Jordan structure (r, r, \dots, r) of d blocks, where d is the nullity. From the description of the centralizer of A given in Proposition 3.1.2, we know B is a $d \times d$ blocked matrix of the form

$$B = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1d} \\ B_{21} & B_{22} & \cdots & B_{2d} \\ \vdots & & & \\ B_{d1} & B_{d2} & \cdots & B_{dd} \end{bmatrix},$$

5. Strictly speaking, the generalized Cayley–Hamilton equation generalizes a *consequence* of the classical Cayley–Hamilton equation, namely, A^n is a linear combination of $I, A, A^2, \dots, A^{n-1}$ for an $n \times n$ matrix A .

6. In terms of the Weyr form, d is the largest Weyr structure component of the various basic λ -Weyr blocks.

where each B_{ij} is an $r \times r$ matrix of the form

$$\begin{bmatrix} a & b & c & \dots & y & z \\ & a & b & c & \dots & y \\ & & a & b & \dots & \\ & & & \ddots & \ddots & \vdots \\ & & & & a & b \\ & & & & & a \end{bmatrix}$$

(the entries a, b, \dots, z will depend on (i, j)). From the discussion in Example 5.1.2, the B_{ij} are just general members of the algebra $R = F[J]$ where J is the $r \times r$ basic nilpotent Jordan matrix. Thus, B can be viewed as a $d \times d$ matrix over the commutative ring R . The coup de grâce of the proof is the observation that the classical Cayley–Hamilton theorem holds not just over fields but over any commutative ring.⁷ Hence, there are polynomials $p_0(J), p_1(J), \dots, p_{d-1}(J)$ in J with coefficients in F such that

$$B^d = p_0(J)I + p_1(J)B + p_2(J)B^2 + \dots + p_{d-1}(J)B^{d-1}.$$

Of course, it is understood here that the “scalars” $p_i(J)$ in this linear combination act under scalar multiplication, which is the same thing as matrix multiplication by the block diagonal matrix $\text{diag}(p_i(J), p_i(J), \dots, p_i(J))$ of d blocks. But the latter matrix is exactly $p_i(A)$ because $A = \text{diag}(J, J, \dots, J)$. Thus,

$$B^d = p_0(A) + p_1(A)B + p_2(A)B^2 + \dots + p_{d-1}(A)B^{d-1}$$

and we are done.

Case (2): A is nilpotent of nonhomogeneous Weyr structure.

We aim to reduce this case to Case (1). By a similarity transformation, we can assume A is a nilpotent Weyr matrix with Weyr structure (n_1, n_2, \dots, n_r) . Note $n_1 = d$. Let $p = dr$ and let $W \in M_p(F)$ be the nilpotent Weyr matrix with the homogeneous Weyr structure (d, d, \dots, d) . Consider the map, defined in top row notation for matrices in the centralizer (see Section 3.4 of Chapter 3),

$$\xi : \mathcal{C}(A) \longrightarrow \mathcal{C}(W), \quad [X_1, X_2, \dots, X_r] \longmapsto [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_r],$$

where the centralizer subalgebras are taken inside $M_n(F)$ and $M_p(F)$, respectively, and where \bar{X}_j is the $d \times d$ matrix $[X_j \ 0]$. That is, each of the $d \times n_j$ blocks X_j has

7. See, for example, Problem 2.4.3 of R. A. Horn and C. R. Johnson’s *Matrix Analysis* or Theorem 7.23 of W. C. Brown’s *Matrices over Commutative Rings*.

been converted to a square matrix \bar{X}_j by appending $d - n_j$ zero columns. In $\xi(X)$, the n columns of a matrix $X \in \mathcal{C}(A)$ occupy new positions, say c_1, c_2, \dots, c_n , but importantly have had only zeros inserted in the (i, c_j) positions for $i \notin \{c_1, \dots, c_n\}$ (each pushing lower column entries down one) to make them $p \times 1$ column vectors. Here, one is making strong use of the description in Proposition 3.2.1 of matrices that centralize a Weyr matrix. For instance, if A has Weyr structure $(3, 2, 1)$, then

$$\xi : \left[\begin{array}{ccc|cc|c} a & b & d & g & i & l \\ 0 & c & e & h & j & m \\ 0 & 0 & f & 0 & k & n \\ \hline & & & a & b & g \\ & & & 0 & c & h \\ & & & & & a \end{array} \right] \mapsto \left[\begin{array}{ccc|ccc|ccc} a & b & d & g & i & 0 & l & 0 & 0 \\ 0 & c & e & h & j & 0 & m & 0 & 0 \\ 0 & 0 & f & 0 & k & 0 & n & 0 & 0 \\ \hline & & & a & b & d & g & i & 0 \\ & & & 0 & c & e & h & j & 0 \\ & & & 0 & 0 & f & 0 & k & 0 \\ \hline & & & & & & a & b & d \\ & & & & & & 0 & c & e \\ & & & & & & 0 & 0 & f \end{array} \right]$$

so $c_i = i$ for $i = 1, 2, \dots, 5$ and $c_6 = 7$. Let \mathcal{S} be the set of all $p \times p$ matrices $M = (m_{ij})$ such that $m_{ij} = 0$ whenever j but not i is in $\{c_1, c_2, \dots, c_n\}$. Then \mathcal{S} is a subalgebra of $M_p(F)$ and the map $\pi : \mathcal{S} \rightarrow M_n(F)$ that deletes the i th row and i th column of a general M for all i in the complementary set of $\{c_1, c_2, \dots, c_n\}$ is an algebra homomorphism. The easiest way to see this is in terms of linear transformations: \mathcal{S} corresponds to the algebra of transformations of F^p that leave invariant the subspace spanned by the standard basis vectors in positions c_1, c_2, \dots, c_n , and π corresponds to the map that restricts those transformations to this subspace. We have $\xi(\mathcal{C}(A)) \subseteq \mathcal{S}$ and π restricted to the image of ξ is the inverse map of ξ . Notice the absence of any claims about ξ being an algebra homomorphism or even $\xi(\mathcal{C}(A))$ being a subalgebra of $M_p(F)$. And for good reasons—both would be false! (See Example 5.2.3 below.) There are some subtleties in this proof.

Having dispensed with those technicalities, we can proceed easily to our goal of the homogeneous reduction. Let $K = \xi(B)$. Since $K \in \mathcal{C}(W)$, we have commuting $p \times p$ matrices W and K in \mathcal{S} such that W is nilpotent with a homogeneous structure and its nullity is still d . By Case (1),

$$K^d = Y_0 + Y_1 K + Y_2 K^2 + \dots + Y_{d-1} K^{d-1}$$

for some matrices $Y_i \in F[W]$. Applying the algebra homomorphism π to this equation, noting $\pi(W) = A$ and $\pi(K) = B$, we obtain our desired result.

Case (3): A is a completely general matrix.

Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be the distinct eigenvalues of A . By the Corollary 1.5.4 to the generalized eigenspace decomposition, and Proposition 3.1.1, we can safely assume that

$$A = \text{diag}(A_1, A_2, \dots, A_k) \text{ and } B = \text{diag}(B_1, B_2, \dots, B_k)$$

with matching block structures, corresponding blocks commuting, and each A_i having λ_i as its sole eigenvalue. To ease notation, we now assume $k = 2$, but the general argument is the same, believe us.

We denote the geometric multiplicities of λ_1 and λ_2 by d and e , respectively, and we can suppose $d \geq e$. Note that any polynomial in $A_1 - \lambda_1 I$ can be expressed as a polynomial in A_1 , and similarly for $A_2 - \lambda_2 I$. Now two applications of the nilpotent version of our equation (established in Cases (1) and (2)), applied to the nilpotent matrices $A_1 - \lambda_1 I$ and $A_2 - \lambda_2 I$, yield

$$\begin{aligned} B_1^d &= s_0(A_1) + s_1(A_1)B_1 + s_2(A_1)B_1^2 + \dots + s_{d-1}(A_1)B_1^{d-1} \\ B_2^d &= t_0(A_2) + t_1(A_2)B_2 + t_2(A_2)B_2^2 + \dots + t_{d-1}(A_2)B_2^{d-1}. \end{aligned}$$

But shouldn't the second equation begin with B_2^e and end with B_2^{e-1} , we hear you ask? Well, yes, but we have modified your equation by multiplying both sides by B_2^{d-e} . Now that you are appeased, let $m_1(x)$ and $m_2(x)$ be the minimal polynomials of A_1 and A_2 , respectively. These have no common factors because they are powers of $x - \lambda_1$ and $x - \lambda_2$, respectively. Consequently, by the Chinese remainder theorem, there are polynomials $p_0(x), p_1(x), \dots, p_{d-1}(x)$ such that

$$\begin{aligned} p_i(x) &\equiv s_i(x) \pmod{m_1(x)} \\ p_i(x) &\equiv t_i(x) \pmod{m_2(x)} \end{aligned}$$

for $i = 0, 1, \dots, d-1$. Observe that, since $m_1(A_1) = 0 = m_2(A_2)$, we have

$$\begin{aligned} B_1^d &= p_0(A_1) + p_1(A_1)B_1 + p_2(A_1)B_1^2 + \dots + p_{d-1}(A_1)B_1^{d-1} \\ B_2^d &= p_0(A_2) + p_1(A_2)B_2 + p_2(A_2)B_2^2 + \dots + p_{d-1}(A_2)B_2^{d-1}. \end{aligned}$$

To finish off, note that, by the simple way in which block diagonal matrices interact algebraically, the common relationship of these two equations carries over to $B = \text{diag}(B_1, B_2)$ and $A = \text{diag}(A_1, A_2)$ to give the desideratum:

$$B^d = p_0(A) + p_1(A)B + p_2(A)B^2 + \dots + p_{d-1}(A)B^{d-1}.$$

□

The Jordan form lends itself well to the proof of Case (1). And in the interest of full disclosure, we must admit that we do not know of a *simple* proof in terms of the Weyr form! When A is nilpotent, homogeneous, and in Weyr form, its Weyr structure is (d, \dots, d) with r terms where $r = n/d$ (and d is the nullity of A). The centralizer description in Proposition 2.3.3 makes B a certain $r \times r$ upper triangular matrix over the *noncommutative* ring $M_d(F)$, a situation not amenable to a classical Cayley–Hamilton argument.

Here is a little example that illustrates the distinction between the classical Cayley–Hamilton equation and the generalized version.

Example 5.2.2

Let

$$A = \begin{bmatrix} 0 & -1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 0 & 1 & -2 \\ 3 & 2 & 2 & -4 \\ -2 & -1 & -1 & 3 \\ 3 & -1 & 0 & -2 \end{bmatrix}.$$

Then A and B commute, and A is nilpotent of nullity 2. (In fact, A has a homogeneous Jordan structure $(2, 2)$.) Therefore, the generalized Cayley–Hamilton equation 5.2.1 guarantees that B satisfies a quadratic equation whose coefficients are polynomials in A . A reworking of the argument in Case (1) gives the equation

$$B^2 = (I - 4A) + (I + 3A)B,$$

as the reader can check. On the other hand, by direct calculation, the characteristic polynomial of B is

$$p(x) = \det(xI - B) = x^4 - 2x^3 - x^2 + 2x + 1,$$

whence the classical Cayley–Hamilton equation satisfied by B is the quartic

$$B^4 = -I - 2B + B^2 + 2B^3.$$

Moreover, one can check that I, B, B^2, B^3 are linearly independent and so $p(x)$ is also the minimal polynomial of B . Therefore, a 4th degree ordinary polynomial equation with coefficients from F is the lowest possible for the vanishing of B . \square

We close this section with an example to show that the mapping ξ used in the proof of the generalized Cayley–Hamilton equation is not an algebra

homomorphism. Had it been, the proof of Case (2) would be much simpler, and that part of the proof would work even for more than two commuting matrices.

Example 5.2.3

Let A be the 3×3 nilpotent Weyr matrix with Weyr structure $(2, 1)$. Let W be the 4×4 nilpotent Weyr matrix with Weyr structure $(2, 2)$. Then

$$\xi : \mathcal{C}(A) \longrightarrow \mathcal{C}(W), \quad \left[\begin{array}{cc|c} a & b & d \\ 0 & c & e \\ \hline & & a \end{array} \right] \mapsto \left[\begin{array}{cc|cc} a & b & d & 0 \\ 0 & c & e & 0 \\ \hline & & a & b \\ & & 0 & c \end{array} \right].$$

Let

$$B = \left[\begin{array}{cc|c} 0 & 0 & 1 \\ 0 & 0 & 0 \\ \hline & & 0 \end{array} \right], \quad C = \left[\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline & & 0 \end{array} \right].$$

Then $B, C \in \mathcal{C}(A)$ but

$$\xi(B)\xi(C) = \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & 0 \\ & & 0 & 0 \end{array} \right] \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & 1 \\ & & 0 & 0 \end{array} \right] = \left[\begin{array}{cc|cc} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & 0 \\ & & 0 & 0 \end{array} \right]$$

whereas $\xi(BC) = \xi(0) = 0$. Thus, ξ is not multiplicative. Note that $\xi(B)\xi(C) \notin \xi(\mathcal{C}(A))$, whence $\xi(\mathcal{C}(A))$ is not even a subalgebra of $M_4(F)$. Note also that B and C commute but $\xi(B)$ and $\xi(C)$ don't commute. That more or less dooms the homogeneous reduction argument used in the proof of Theorem 5.2.1 for the case of 3-generated commutative subalgebras $F[A, B, C]$. (That is, in Case (2), our argument won't produce ξ such that $\xi(A), \xi(B), \xi(C)$ commute and $\xi(A)$ is nilpotent with a homogeneous structure.) \square

5.3 PROOF OF GERSTENHABER'S THEOREM

Now to the third and main part of the Barría–Halmos argument for establishing Gerstenhaber's theorem, the inductive step. It is here that the Weyr form suggests a great simplification to their argument, that was formulated in terms of the Jordan form.

Theorem 5.3.1

Let A and B be commuting nilpotent $n \times n$ matrices over a field F , and let (n_1, n_2, \dots, n_r) be the Weyr structure of A . Then the following collection \mathcal{B} of n matrices spans (as a vector space) the subalgebra $\mathcal{A} = F[A, B]$ of $M_n(F)$ generated by A and B :

$$\begin{array}{ccccccc} I, & B, & B^2, & \dots, & B^{n_1-1} \\ A, & BA, & B^2A, & \dots, & B^{n_2-1}A \\ A^2, & BA^2, & B^2A^2, & \dots, & B^{n_3-1}A^2 \\ & \vdots & & & \\ & \vdots & & & \\ & \vdots & & & \\ A^{r-1}, & BA^{r-1}, & B^2A^{r-1}, & \dots, & B^{n_r-1}A^{r-1}. \end{array}$$

In particular, $\dim \mathcal{A} \leq n$.

Proof

We proceed by induction on r (which is the index of nilpotency of A). If $r = 1$, then $A = 0$ and $n_1 = n$ so the displayed spanning set (only the first row is nonzero) comes directly from the classical Cayley–Hamilton equation applied to B .

Now suppose $r > 1$. We can assume A is already in Weyr form because the theorem will hold for A and B if it holds for $C^{-1}AC$ and $C^{-1}BC$ for some invertible matrix C . Let \mathcal{T} be the subalgebra of $M_n(F)$ consisting of all block upper triangular matrices with the same block structure as A , that is, with diagonal blocks of size n_1, n_2, \dots, n_r . Certainly $\mathcal{A} \subseteq \mathcal{T}$ by Proposition 2.3.3. Consider the projection

$$\pi : \mathcal{A} \rightarrow \mathcal{T}$$

of \mathcal{A} onto its bottom right $(r-1) \times (r-1)$ corner of blocks:

$$\begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1r} \\ 0 & X_{22} & X_{23} & \cdots & X_{2r} \\ 0 & 0 & X_{33} & \cdots & X_{3r} \\ \vdots & & & & \\ 0 & 0 & \cdots & 0 & X_{rr} \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & X_{22} & X_{23} & \cdots & X_{2r} \\ 0 & 0 & X_{33} & \cdots & X_{3r} \\ \vdots & & & & \\ 0 & 0 & \cdots & 0 & X_{rr} \end{bmatrix}$$

This is a (very natural) algebra homomorphism (although it does not preserve the identity). Also, $\pi(\mathcal{A})$ sits naturally inside the algebra of $(n - n_1) \times (n - n_1)$ matrices over F . (To see this, apply the bottom right corner version of the projection η of Proposition 1.2.1 for $m = n_1$. When restricted to $\pi(\mathcal{A})$, this η is a 1-1 algebra homomorphism.) Viewed inside $M_{n-n_1}(F)$, the matrix $\pi(A)$ is still in Weyr form with Weyr structure (n_2, n_3, \dots, n_r) . So we are nicely set up for an inductive step — apply the theorem to the commuting nilpotent matrices $\pi(A)$ (of index

$r - 1$) and $\pi(B)$. Since π is an algebra homomorphism, we have by induction on the nilpotency index that $\pi(\mathcal{A}) = F[\pi(A), \pi(B)]$ is spanned (as a vector space) by $\pi(\mathcal{B}')$, where \mathcal{B}' consists of

$$\begin{array}{ccccccc} I, & B, & B^2, & \dots, & B^{n_2-1} \\ A, & BA, & B^2A, & \dots, & B^{n_3-1}A \\ A^2, & BA^2, & B^2A^2, & \dots, & B^{n_4-1}A^2 \\ & \vdots & & & \\ & \vdots & & & \\ & \vdots & & & \\ A^{r-2}, & BA^{r-2}, & B^2A^{r-2}, & \dots, & B^{n_r-1}A^{r-2}. \end{array}$$

Claim: $\ker(\pi) = \text{ann}(A)$, where $\text{ann}(A) = \{X \in \mathcal{A} : XA = 0\}$ is the annihilator ideal of A (within \mathcal{A}).

To justify the claim, observe that $X \in \mathcal{A}$ lies in $\ker(\pi)$ exactly when, as a blocked matrix, its rows 2, 3, \dots , r are zero. However, because X centralizes A , Proposition 2.3.3 stipulates that the first row of blocks of X must take the form

$$[X_{11}, X_{12}, \dots, X_{1r}],$$

where for $j = 1, \dots, r - 1$, X_{1j} is an $n_1 \times n_j$ matrix with its first n_{j+1} columns all zero, because of the form

$$X_{1j} = \begin{bmatrix} X_{2,j+1} & * \\ 0 & * \end{bmatrix}.$$

But from the shifting and partial deleting of rightmost columns that right multiplication by A performs on the blocks of X (see Remark 2.3.1), we see that these are precisely the conditions for X to lie in $\text{ann}(A)$.

Let $X \in \mathcal{A}$. Since $\pi(X)$ is in the span of $\pi(\mathcal{B}')$, we can write $\pi(X) = \pi(Y)$ for some Y in the span of \mathcal{B}' . By our above claim, $X - Y \in \text{ann}(A)$, whence $XA = YA$. Therefore, since YA is in the span $\langle \mathcal{B}'A \rangle$ of $\mathcal{B}'A$, we have that

$$(*) \quad XA \text{ is in the span of } \mathcal{B}'A \text{ for all } X \in \mathcal{A} = F[A, B].$$

In other words, *any* product in $F[A, B]$ having A as a factor is automatically in $\langle \mathcal{B}'A \rangle$, and therefore is in the span $\langle \mathcal{B} \rangle$ of \mathcal{B} because $\mathcal{B}'A \subseteq \mathcal{B}$. Next observe that n_1 is the nullity of A , because it is the size of the first block of the Weyr form. Moreover,

$$\mathcal{B} = \{I, B, B^2, \dots, B^{n_1-1}\} \cup \mathcal{B}'A.$$

Now the generalized Cayley–Hamilton Equation 5.2.1 (with $d = n_1$) shows that $\langle \mathcal{B} \rangle$ contains all powers of B , whence, by $(*)$, $\langle \mathcal{B} \rangle$ is invariant under multiplication

by B . And $(*)$ shows that $\langle B \rangle$ is also invariant under multiplication by A . Moreover, since $\langle B \rangle$ clearly contains I , A , and B , this shows $F[A, B]$ is contained in $\langle B \rangle$. The reverse containment obviously holds. Therefore, $F[A, B]$ is spanned by B , as desired. This completes the induction and the proof of the theorem. \square

Having completed our assigned three steps (Proposition 5.1.1 for $k = 2$, Theorem 5.2.1, and Theorem 5.3.1) for establishing Gerstenhaber's theorem, we can now formally state the result:⁸

Theorem 5.3.2 (Gerstenhaber)

If A and B are commuting $n \times n$ matrices over any field F , then the subalgebra they generate has dimension at most n .

Remark 5.3.3

As a flipside to Gerstenhaber's theorem, a theorem of Burnside⁹ says that over an algebraically closed field F , if \mathcal{A} is a subalgebra of $M_n(F)$ that does not leave invariant (under left multiplication) any proper subspace of F^n , then $\mathcal{A} = M_n(F)$. In particular, $\dim \mathcal{A} = n^2$, the opposite extreme to Gerstenhaber's result. Of course, if A and B are commuting $n \times n$ matrices over F , the subalgebra $F[A, B]$ leaves invariant even some 1-dimensional subspaces of F^n , namely, any subspace spanned by a common eigenvector of A and B . One might therefore expect that sandwiched between the theorems of Gerstenhaber and Burnside is a result saying that if A and B are (noncommuting) $n \times n$ matrices that have no common eigenvector, then $n < \dim F[A, B] \leq n^2$. Not so. One can quickly check that the following pair of 4×4 matrices have no common eigenvector, yet generate only a 4-dimensional subalgebra:

$$A = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array} \right], \quad B = \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right]. \quad \square$$

By the duality between the Jordan and Weyr forms, our spanning set result for $F[A, B]$ in Theorem 5.3.1 has a dual. Transposing the spanning set, by writing its columns as rows, and using the fact that Jordan and Weyr structures are dual partitions (Theorem 2.4.1), we obtain the original statement of Barría and Halmos (and that too of Laffey and Lazarus). Simply observe that transposing the spanning set mirrors transposing the Young

8. Our statement of Gerstenhaber's theorem is the common one, but in fact Gerstenhaber proved the stronger result that any 2-generated commutative subalgebra of $M_n(F)$ is contained in some n -dimensional commutative subalgebra.

9. See the 1998 article by T. Y. Lam.

tableau, and the latter determines the indices of the new array for the same spanning set.¹⁰

Corollary 5.3.4

Let A and B be commuting nilpotent $n \times n$ matrices over a field F , and let (m_1, m_2, \dots, m_s) be the Jordan structure of A . Then the following collection \mathcal{B} of n matrices spans the subalgebra $\mathcal{A} = F[A, B]$ of $M_n(F)$ generated by A and B :

$$\begin{array}{ccccccc} I, & A, & A^2, & \dots, & A^{m_1-1} \\ B, & BA, & BA^2, & \dots, & BA^{m_2-1} \\ B^2, & B^2A, & B^2A^2, & \dots, & B^2A^{m_3-1} \\ & \cdot & & & \\ & \cdot & & & \\ & \cdot & & & \\ B^{s-1}, & B^{s-1}A, & B^{s-1}A^2, & \dots, & B^{s-1}A^{m_s-1}. \end{array}$$

In their proof of this dual of our Theorem 5.3.1, with A now in Jordan form and with Jordan structure (m_1, m_2, \dots, m_s) , Barría and Halmos also used projection as their inductive wedge, in their case onto the top left-hand $(s-1) \times (s-1)$ corner of blocks of members of $F[A, B]$. *But projection in this setting is not a multiplicative map for the Jordan form!* However, with the benefit of hindsight, and retracking using our duality connection (Theorem 2.4.1), the algebra homomorphism analogous to the one we have used for the Weyr form is to block the matrices according to the Jordan structure and then delete the first row and first column from each of the s^2 blocks of matrices in $F[A, B]$. The inductive step then proceeds pretty much the same as above (now $\pi(A)$ has Jordan structure $(m_1-1, m_2-1, \dots, m_s-1)$, but of course we need to drop the terms m_i-1 when $m_i=1$). We leave it to the interested reader to check out the details.

As regards our own projection map used in Theorem 5.3.1, it would not have been unreasonable to try instead to project onto the *top* $(r-1) \times (r-1)$ left-hand corner of blocks. This is still an algebra homomorphism but unfortunately it has the wrong kernel for the induction to work. However, there are other situations (some in Section 5.5) where this is indeed the correct projection to use.

10. The quickest way to see this is to look at a simple example, say when A has Weyr structure $(5, 3, 2)$. The transposed spanning set then has the shape of the dual partition $(3, 3, 2, 1, 1)$, which corresponds to the Jordan structure of A .

We remark in closing this section that it is an open problem as to whether the conclusion of Gerstenhaber's theorem holds for three commuting matrices. However, there are easy examples to show it fails for four or more. In fact, as we show in Example 6.3.4 of Chapter 6, for each $n \geq 4$, there exist four commuting $n \times n$ matrices A_1, A_2, A_3, A_4 for which $\dim F[A_1, A_2, A_3, A_4] = n + 1$.

5.4 MAXIMAL COMMUTATIVE SUBALGEBRAS

Gerstenhaber's theorem invites the following question: when is a 2-generated commutative subalgebra $F[A, B]$ of $M_n(F)$ a **maximal commutative subalgebra**? Here, maximal means $F[A, B]$ is not properly contained in any other commutative subalgebra \mathcal{C} of $M_n(F)$ (where \mathcal{C} need not be 2-generated). We approach this question using the Weyr form, and particularly the leading edge subspaces, which were introduced in Section 3.4 of Chapter 3. In the process we give new derivations of some known results.

To conform with the earlier definition in Chapter 3 of the centralizer $\mathcal{C}(A)$ of a matrix A , we can define the **centralizer** $\mathcal{C}(\mathcal{A})$ of a **subalgebra** \mathcal{A} of $M_n(F)$ to be the set of all matrices that commute with every matrix in \mathcal{A} . (The connection is $\mathcal{C}(A) = \mathcal{C}(F[A])$.) A **self-centralizing subalgebra** of $M_n(F)$ is a subalgebra \mathcal{A} for which $\mathcal{C}(\mathcal{A}) = \mathcal{A}$. In these terms, a self-centralizing subalgebra \mathcal{A} is exactly a maximal commutative subalgebra. Certainly the forward implication holds. The converse also holds because if \mathcal{A} is commutative and $X \in \mathcal{C}(\mathcal{A})$ but $X \notin \mathcal{A}$, then the subalgebra generated by \mathcal{A} and X is commutative and strictly contains \mathcal{A} .

In the early 1990s, Laffey and Lazarus, and Neubauer and Saltman, independently characterized the 2-generated commutative subalgebras $F[A, B]$ of $M_n(F)$ that are maximal. They are exactly the 2-generated commutative subalgebras of dimension n .¹¹ By Proposition 5.1.1, the proof reduces to the core case where A and B are nilpotent. We will give a new proof of the characterization in the case when A also has a homogenous Weyr structure. We begin with a little result that can be viewed as a strengthened version of Gerstenhaber's theorem in the homogeneous case.

Lemma 5.4.1

Let A and B be commuting $n \times n$ matrices over F and assume A is a nilpotent Weyr matrix with a homogeneous Weyr structure (d, d, \dots, d) of r blocks (so $n = dr$).

11. Without the 2-generated assumption, maximal commutative subalgebras can even have dimension less than n . In 1965, R. C. Courter constructed a maximal commutative subalgebra of $M_{14}(F)$ of dimension 13. We say a little more on this in Section 6.6 of Chapter 6.

Let $\mathcal{A} = F[A, B]$ and let U_0, U_1, \dots, U_{r-1} be the leading edge subspaces of \mathcal{A} relative to A . Then $\dim U_i \leq d$ for all i .

Proof

Before we embark on the proof proper, we note that if \mathcal{A} is a commutative subalgebra of $M_n(F)$ containing A and the leading edge subspaces of \mathcal{A} satisfy $\dim U_i \leq d$ for all i , then by Theorem 3.4.3,

$$\dim \mathcal{A} = \dim U_0 + \dim U_1 + \dots + \dim U_{r-1} \leq d + d + \dots + d = dr = n.$$

Hence, $\dim \mathcal{A} \leq n$. It is in this sense that the conclusion of the lemma is a strengthened version of Gerstenhaber's theorem. However, we can't derive the theorem from the lemma because in fact we use Gerstenhaber's theorem in the proof of the lemma.

Now let $\mathcal{A} = F[A, B]$ be as in the lemma. By Proposition 3.4.4(3), we know $\dim U_0 \leq \dim U_1 \leq \dots \leq \dim U_{r-1}$ and so it is enough to show $\dim U_{r-1} \leq d$. To the contrary, suppose $\dim U_{r-1} = e > d$. Choose a positive integer $s > r$ such that

$$(*) \quad \sum_{i=0}^{r-1} \dim U_i + (s-r)e > ds$$

that is, $s > (re - \sum_{i=0}^{r-1} \dim U_i)/(e - d)$. Let \bar{A} be the $ds \times ds$ nilpotent Weyr matrix with Weyr structure (d, d, \dots, d) (an s -tuple), and extend B to a $ds \times ds$ matrix $\bar{B} \in \mathcal{C}(\bar{A})$. For instance, in top row notation, if $B = [B_0, B_1, \dots, B_{r-1}]$, take $\bar{B} = [B_0, B_1, \dots, B_{r-1}, 0, 0, \dots, 0]$. (This step is usually not possible in the nonhomogeneous case—a subtle point.) Let $\bar{\mathcal{A}} = F[\bar{A}, \bar{B}]$. Let $\bar{U}_0, \bar{U}_1, \dots, \bar{U}_{s-1}$ be the leading edge subspaces of $\bar{\mathcal{A}}$ relative to \bar{A} . Observe from the block upper triangular matrices involved (or by a projection homomorphism onto an appropriate corner of blocks) that for $i = 0, 1, \dots, r-1$ the leading edge subspaces of $\bar{\mathcal{A}}$ are the same as those for \mathcal{A} , namely U_0, U_1, \dots, U_{r-1} . In particular, $\dim \bar{U}_i \geq \dim U_{r-1}$ for $i = r, \dots, s-1$ by Proposition 3.4.4(3). Now we compute the dimension of $\bar{\mathcal{A}}$ as the sum of its leading edge dimensions (Theorem 3.4.3):

$$\begin{aligned} \dim \bar{\mathcal{A}} &= \sum_{i=0}^{s-1} \dim \bar{U}_i \\ &= \sum_{i=0}^{r-1} \dim \bar{U}_i + \sum_{i=r}^{s-1} \dim \bar{U}_i \\ &\geq \sum_{i=0}^{r-1} \dim U_i + (s-r) \dim U_{r-1} \end{aligned}$$

$$\begin{aligned}
&= \sum_0^{r-1} \dim U_i + (s-r)e \\
&> ds \text{ by } (*)
\end{aligned}$$

This contradicts Gerstenhaber's Theorem 5.3.2 for the dimension of 2-generated commutative subalgebras of $M_{ds}(F)$. Therefore, we must have $\dim U_{r-1} \leq d$. \square

Example 5.4.2

Lemma 5.4.1 fails in the nonhomogeneous case. For instance, the reader can check that using a $(2, 1)$ Weyr structure and taking

$$A = \left[\begin{array}{cc|c} 0 & 0 & 1 \\ 0 & 0 & 0 \\ \hline & & 0 \end{array} \right], \quad B = \left[\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & 1 \\ \hline & & 0 \end{array} \right]$$

produces $\dim U_0 = 1$, $\dim U_1 = 2$. So it is not true in general that $\dim U_i \leq n_{i+1}$ if the Weyr structure is (n_1, n_2, \dots, n_r) .

The lemma also fails for 3-generated commutative subalgebras $F[A, B, C]$ with A a homogeneous nilpotent Weyr matrix. One can check that using a $(2, 2)$ Weyr structure and taking

$$A = \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline & & 0 & 0 \\ & & 0 & 0 \end{array} \right], \quad B = \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & 0 \\ & & 0 & 0 \end{array} \right], \quad C = \left[\begin{array}{cc|cc} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & 0 \\ & & 0 & 0 \end{array} \right]$$

produces $\dim U_0 = 1$, $\dim U_1 = 3$. \square

Our next theorem can be viewed as the Weyr form dual of a result established by Laffey and Lazarus in 1991 but with a very much shorter proof, and no restriction on the characteristic of F . The theorem also appears as an equivalent result, in terms of a homogeneous Jordan matrix (and an abstract view of its centralizer in terms of polynomials) in the 2009 paper by Sethuraman and Šivic(their Theorem 2.4). Recall that a nonderogatory matrix is one whose eigenspaces are 1-dimensional (see Section 1.1 of Chapter 1 and various characterizations in Proposition 3.2.4).

Theorem 5.4.3

Let A and B be commuting $n \times n$ matrices over F and assume A is a nilpotent Weyr matrix with a homogeneous Weyr structure (d, d, \dots, d) of r blocks (so $n = dr$).

As an $r \times r$ blocked matrix of $d \times d$ blocks, let

$$B = \begin{bmatrix} B_0 & B_1 & B_2 & \cdots & B_{r-2} & B_{r-1} \\ & B_0 & B_1 & B_2 & & B_{r-2} \\ & & B_0 & B_1 & \ddots & \vdots \\ & & & \ddots & \ddots & B_2 \\ & & & & B_0 & B_1 \\ & & & & & B_0 \end{bmatrix}.$$

Then $\dim F[A, B] = n$ if and only if B_0 is a nonderogatory $d \times d$ matrix.

Proof

Let U_0, U_1, \dots, U_{r-1} be the leading edge subspaces of $\mathcal{A} = F[A, B]$ relative to A . Note that $U_0 = F[B_0]$. From our earlier work, we are in possession of three important pieces of information:

- (1) $\dim \mathcal{A} = \dim U_0 + \dim U_1 + \cdots + \dim U_{r-1}$ (Theorem 3.4.3)
- (2) $\dim U_0 \leq \dim U_1 \leq \cdots \leq \dim U_{r-1}$ (Proposition 3.4.4(3))
- (3) $\dim U_{r-1} \leq d$ (Lemma 5.4.1)

Hence,

$$\begin{aligned} \dim \mathcal{A} = n (= dr) &\iff \dim U_0 = d \\ &\iff \dim F[B_0] = d \\ &\iff B_0 \text{ is nonderogatory.} \quad \square \end{aligned}$$

Theorem 5.4.3 enables us to give a relatively simple proof in the homogeneous case of the Laffey and Lazarus, and Neubauer and Saltman, characterization of 2-generated self-centralizing subalgebras. The proof of the full theorem would take us too far afield. Our goal is the illustration of various applications of the Weyr form, not the study of commutative subalgebras of matrices per se.

Theorem 5.4.4 (Laffey–Lazarus; Neubauer–Saltman)

Let A and B be commuting $n \times n$ matrices over an algebraically closed field F . Then $F[A, B]$ is a self-centralizing subalgebra of $M_n(F)$ (equivalently, a maximal commutative subalgebra) if and only if $\dim F[A, B] = n$.

Proof

By Proposition 5.1.1, we can reduce to the case where A is nilpotent. We will consider only the case where A has a homogeneous Weyr structure (d, d, \dots, d) , with, say, r blocks (so $n = dr$). Suppose $\dim F[A, B] = n$. After a similarity transformation, we can assume A and B have the form described in Theorem 5.4.3 with B_0 a nonderogatory $d \times d$ matrix. Let $\mathcal{A} = F[A, B]$ and let U_0, U_1, \dots, U_{r-1} be the leading edge subspaces of \mathcal{A} relative to A . Now let $C \in \mathcal{C}(\mathcal{A})$ and let $\overline{\mathcal{A}} = F[A, B, C]$ be the commutative subalgebra of $M_n(F)$ generated by A, B, C . Let $\overline{U}_0, \overline{U}_1, \dots, \overline{U}_{r-1}$ be the leading edge subspaces of $\overline{\mathcal{A}}$ (again relative to the Weyr matrix A). Clearly $U_i \subseteq \overline{U}_i$ for all i . By Proposition 3.4.4 (3), $U_0 \subseteq \overline{U}_i$ for all i . Note that since $U_0 = F[B_0]$ and B_0 is nonderogatory, U_0 is a self-centralizing subalgebra of $M_d(F)$ (see Proposition 3.2.4). But by Proposition 3.4.4 (4), U_0 centralizes each of the \overline{U}_i , whence we must have $\overline{U}_i = U_0$ for all i . In particular $\overline{U}_i = U_i$ (because $U_0 \subseteq U_i \subseteq \overline{U}_i$) and so $\dim \overline{\mathcal{A}} = \dim \mathcal{A}$ by Theorem 3.4.3. Inasmuch as \mathcal{A} is a subalgebra of $\overline{\mathcal{A}}$, this implies $\mathcal{A} = \overline{\mathcal{A}}$. Therefore $C \in \mathcal{A}$, which shows that \mathcal{A} is a self-centralizing subalgebra.

For the converse, suppose $\mathcal{A} = F[A, B]$ is a self-centralizing subalgebra of $M_n(F)$. We can assume A and B have the form described in Theorem 5.4.3, and we need only show B_0 is a nonderogatory $d \times d$ matrix to conclude that $\dim \mathcal{A} = n$. For each $X \in \mathcal{C}(B_0)$, we have in top row notation the matrix $[0, 0, \dots, 0, X]$ which commutes with both A and B , and consequently commutes with everything in $\mathcal{A} = F[A, B]$. Therefore since \mathcal{A} is self-centralizing, we must have $[0, 0, \dots, 0, X] \in \mathcal{A}$. This places X in the leading edge subspace U_{r-1} of \mathcal{A} . Thus, $\mathcal{C}(B_0) \subseteq U_{r-1}$. But we know from Lemma 5.4.1 that $\dim U_{r-1} \leq d$. Hence, $\dim \mathcal{C}(B_0) \leq d$. On the other hand, by the Frobenius formula 3.1.3 and the standard nilpotent reduction argument of Proposition 3.1.1, we know $\dim \mathcal{C}(B_0) \geq d$ whence $\dim \mathcal{C}(B_0) = d$. (See the argument in the proof of Proposition 3.2.4.) By Proposition 3.2.4, this in turn implies B_0 is nonderogatory. We are finished. \square

Example 5.4.5

At the risk of appearing to use a cannon to shoot a sparrow, let us illustrate the Gerstenhaber, Laffey–Lazarus, Neubauer–Saltman theorems by showing that all commutative subalgebras \mathcal{A} of $M_3(F)$ have dimension at most 3. If $\dim \mathcal{A} > 3$ we could choose linearly independent matrices $I, X, Y, Z \in \mathcal{A}$ (where I is the identity matrix). Then $F[X, Y]$ is a 2-generated commutative subalgebra of $M_3(F)$ of dimension at least 3 (it contains I, X, Y) and so its dimension must be exactly 3 by Gerstenhaber's Theorem 5.3.2. Now by the Laffey–Lazarus, Neubauer–Saltman Theorem 5.4.4, $F[X, Y]$ must be self-centralizing in $M_3(F)$, which is contradicted by $Z \notin F[X, Y]$ (note that $\{I, X, Y\}$ is a basis for $F[X, Y]$ here). Therefore, $\dim \mathcal{A} \leq 3$. \square

Finally, we note that Theorem 5.4.4 can fail for three commuting matrices $A, B, C \in M_n(F)$: having $\dim F[A, B, C] = n$ doesn't guarantee that $F[A, B, C]$ is self-centralizing. In terms of matrix units e_{ij} (zero entries except for a 1 in the (i, j) position), the matrices e_{13}, e_{14}, e_{24} generate a 4-dimensional commutative subalgebra of $M_4(F)$ (we could even replace e_{13} by the homogeneous nilpotent matrix $e_{13} + e_{24}$). However, this subalgebra is not a maximal commutative subalgebra because $e_{13}, e_{14}, e_{24}, e_{23}$ generate an even larger commutative subalgebra of dimension 5 (see Example 6.3.4). So $F[e_{13}, e_{14}, e_{24}]$ is not self-centralizing. Notice how our argument in Example 5.4.5 breaks down for showing commutative subalgebras of $M_4(F)$ have dimension at most 4.

5.5 PULLBACKS AND 3-GENERATED COMMUTATIVE SUBALGEBRAS

Just about every technique we have used in studying Gerstenhaber's theorem and self-centralizing subalgebras, in the case of 2-generated commutative subalgebras of $M_n(F)$, fails for three generators (even in the homogeneous case). And this is despite all known triples of commuting $n \times n$ matrices satisfying the conclusion of Gerstenhaber's theorem—the subalgebra they generate has dimension at most n .¹² The goal of this section is to show how the Weyr form suggests new techniques for establishing n as an upper bound in the case of three commuting generators. These techniques do at least work in low order cases. Again, since it is not our objective to study commutative subalgebras for their own sake, but rather to illustrate the utility of the Weyr form, we will try to keep the illustrations relatively simple.¹³ Nevertheless, at times the material in this section may seem a little technical, requiring the concentration of a kookaburra.¹⁴ However, the reader should feel free to just skim (or skip) the details, although we do suggest that he or she at least check out the concept of a pullback and how it opens up the possibility of an inductive argument. Of course, we hope those readers who are interested in extensions of Gerstenhaber's theorem will study every detail. As far as we know, the tools developed here have not appeared elsewhere in the literature.

12. As a betting man, the first author is willing to wager that there do exist three commuting matrices for which this dimension is greater than n .

13. This is not always easy to do because the underlying questions are inherently very difficult.

14. Kookaburras (four species) are large birds of the kingfisher family, native to Australia and New Guinea. In a group, they often break into unmistakably loud, hysterical, human-like laughter. A kookaburra can sit very still for long periods, before swooping on its unsuspecting prey (grubs, frogs, mice, snakes), up to 15 or 20 meters away.

We will work with three commuting matrices W, K, K' . By the standard reduction, we can assume they are nilpotent with W in Weyr form. We will also assume that W has a homogeneous Weyr structure. Let us fix some notation for the remainder of the section.

Notation.

$W =$ an $n \times n$ nilpotent Weyr matrix with Weyr structure (d, d, \dots, d) .

$r =$ # blocks in the Weyr structure (so $n = dr$).

$\mathcal{A} = F[W, K, K']$, a 3-generated commutative subalgebra of $M_n(F)$.

$K = [D_0, D_1, \dots, D_{r-1}]$ as an $n \times n$ matrix in top row notation (relative to W).

$K' = [D'_0, D'_1, \dots, D'_{r-1}]$ as an $n \times n$ matrix in top row notation.

$\mathcal{T} =$ the algebra of $r \times r$ block upper triangular matrices with $d \times d$ blocks.

U_0, U_1, \dots, U_{r-1} are the leading edge subspaces of \mathcal{A} relative to W .

Our choice of letters has been guided by “ W ” for “Weyr,” “ K ” for “commuting,”¹⁵ and “ D ” for diagonal (the latter will become clearer in Chapter 6). The primes ‘ provide symmetry and economy. To remind the reader what top row notation (introduced in Section 4 of Chapter 3) is short for, as a full $r \times r$ blocked matrix with $d \times d$ blocks,

$$K = \begin{bmatrix} D_0 & D_1 & D_2 & \dots & D_{r-2} & D_{r-1} \\ & D_0 & D_1 & D_2 & \ddots & D_{r-2} \\ & & D_0 & D_1 & \ddots & \vdots \\ & & & \ddots & \ddots & D_2 \\ & & & & D_0 & D_1 \\ & & & & & D_0 \end{bmatrix}.$$

We next introduce the concept of a “pullback” of a matrix that centralizes W . (With suitable modifications, the concept can sometimes be used in the nonhomogeneous case as well, but we shall not pursue that here.)

Definition 5.5.1: Using top row notation, let $B = [B_0, B_1, \dots, B_{r-1}] \in \mathcal{C}(W)$ with $B_0 = B_1 = \dots = B_{g-1} = 0$ for some positive integer g . A **g -block pullback** of B is

15. Think of the German “ k ” for “kommutativ.”

any $X \in \mathcal{C}(W)$ with $XW^g = B$. We also define a **0-block pullback** of B to be simply $X = B$ (and so we still have $XW^0 = B$).

What do pullbacks look like? In top row notation, a g -block pullback of B is simply any $X \in \mathcal{C}(W)$ of the form $X = [B_g, B_{g+1}, \dots, B_{r-1}, *, *, \dots, *]$ where the $*$ entries are arbitrary. This is because under right multiplication by W , each block is shifted one step to the right, the last block is annihilated, and a zero first block is introduced. (See Remark 2.3.1.) Thus, the blocks of B have been “pulled back g steps.” For instance, if W has Weyr structure $(3, 3, 3)$ and

$$B = \left[\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 3 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 9 & 1 & 4 \end{array} \right],$$

then

$$X = \left[\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 1 & 3 & 7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 9 & 1 & 4 & 0 & 0 & 0 \end{array} \right] \text{ and } Y = \left[\begin{array}{ccc|ccc|ccc} 1 & 3 & 7 & 0 & 0 & 1 & 0 & 0 & 0 \\ 6 & 2 & 2 & 0 & 3 & 0 & 0 & 0 & 2 \\ 9 & 1 & 4 & -5 & 0 & 0 & 0 & 1 & 6 \end{array} \right]$$

are 1-block and 2-block pullbacks of B , respectively. Pullbacks could also be formulated in terms of the Jordan form but they would become unwieldy.

Lemma 5.5.2

Suppose g, h are integers between 0 and r such that $D_0 = D_1 = \dots = D_{g-1} = 0$ and $D'_0 = D'_1 = \dots = D'_{h-1} = 0$. Let X be a g -block pullback of K and let X' be an h -block pullback of K' . Then:

- (1) $(X'X - XX')W^{g+h} = 0$,
- (2) $K'K = (X'X)W^{g+h}$,
- (3) $KK' = (XX')W^{g+h}$,
- (4) $K^2 = X^2W^{2g}$,
- (5) $(K')^2 = (X')^2W^{2h}$.

Proof

Our hypotheses imply $K = XW^g$ and $K' = X'W^h$. For (1) we argue that because W, K , and K' commute,

$$\begin{aligned}
 (X'X - XX')W^{g+h} &= X'XW^gW^h - XX'W^hW^g \\
 &= X'(KW^h) - X(K'W^g) \\
 &= X'(W^hK) - X(W^gK') \\
 &= K'K - KK' = 0.
 \end{aligned}$$

For (2) we have

$$\begin{aligned}
 K'K &= X'W^hK \\
 &= X'KW^h \\
 &= X'XW^gW^h \\
 &= X'XW^{g+h}.
 \end{aligned}$$

Clearly (3), (4), and (5) are just special cases of (2). □

In general, the pullbacks X and X' in the proposition won't commute. So how can they be of use to us? Property (1) says that X and X' commute modulo being right annihilated by W^{g+h} , so we should be aiming to divide out by the left annihilator ideal of W^{g+h} inside the ring $\mathcal{C}(W)$. The correct homomorphism will do this for us. In fact, the appropriate homomorphism is a corner projection map yet again.

Proposition 5.5.3

Under the same hypotheses as the previous lemma, let $m = g + h$ and suppose $m < r$. Let

$$\pi : \mathcal{T} \longrightarrow \mathcal{T}$$

be the projection map onto the $(r - m) \times (r - m)$ top left corner of blocks of matrices of \mathcal{T} (regarding members of \mathcal{T} as $r \times r$ blocked matrices with $d \times d$ blocks). Then:

- (1) π is an algebra homomorphism.
- (2) The kernel of π is the left annihilator ideal of W^m , that is, $\ker \pi = \{X \in \mathcal{T} : XW^m = 0\}$.
- (3) The matrices $\pi(W), \pi(X), \pi(X')$ commute.

- (4) $F[\pi(W), \pi(X), \pi(X')]$ can be regarded as a 3-generated commutative subalgebra of $M_p(F)$, for $p = n - md$, in which $\pi(W)$ is a nilpotent Weyr matrix of Weyr structure (d, d, \dots, d) with $r - m$ blocks.

Proof

(1) is routine. From the shifting effect under right multiplication by W on the columns of an $r \times r$ (blocked) matrix in \mathcal{T} , we see W^m right annihilates matrices whose first $r - m$ columns are zero. But these are precisely the matrices in the kernel of π . This gives (2). It is clear that viewed as a $p \times p$ corner matrix, $\pi(W)$ is a nilpotent Weyr matrix whose Weyr structure is that of W after the last m terms in its structure have been removed. Moreover, in the same corner, both $\pi(X)$ and $\pi(X')$ have the correct form to centralize $\pi(W)$ (Proposition 2.3.3). From Lemma 5.5.2 (1), $X'X - XX'$ left annihilates W^m , whence by (2) we have $X'X - XX' \in \ker \pi$. Therefore by (1), $\pi(X')\pi(X) - \pi(X)\pi(X') = \pi(X'X - XX') = 0$, which says that $\pi(X)$ and $\pi(X')$ commute. This establishes (3) and (4). \square

As the reader may have guessed, we are aiming to use pullbacks and the projection in Proposition 5.5.3 to initiate some sort of inductive argument. Before we can do that, we need one more little result.

Lemma 5.5.4

For the projection $\pi : \mathcal{T} \rightarrow \mathcal{T}$ in Proposition 5.5.3, the following hold:

- (1) Each $X \in \mathcal{T}$ can be written as $X = \pi(X) + Y$ where $Y \in \mathcal{T}$ satisfies $YW^m = 0$.
- (2) Each $X \in \mathcal{T}$ satisfies $\pi(X)W^m = XW^m$.

Proof

By the very nature of a projection map, $X - \pi(X) \in \ker \pi$. But from Proposition 5.5.3 (2), we know W^m right annihilates kernel members. Thus, we have (1), and of course (2) follows immediately from (1). \square

When $d = 1$ (that is, W has Weyr structure $(1, 1, \dots, 1)$), W is a 1-regular matrix and as we noted in Example 5.1.2, \mathcal{A} is then just $F[W]$ and has dimension n . As a corollary to our approximate simultaneous diagonalization results in Chapter 6, we will see that over the complex field, $\dim \mathcal{A} \leq n$ when $d = 2$. (The result holds also over a general field.) We make the next step in deciding if Gerstenhaber's theorem holds for three commuting matrices.

Assumption: For the remainder of this section we assume W is (strictly) 3-regular, that is, $d = 3$. Hence, W is an $n \times n$ nilpotent Weyr matrix having Weyr structure $(3, 3, \dots, 3)$, with $r = n/3$ blocks.

In view of Gerstenhaber's theorem, we can make further simplifications as in our next lemma.

Lemma 5.5.5

Without loss of generality (when W is 3-regular), to establish the bound $\dim F[W, K, K'] \leq n$ we can assume that

$$K = [D_0, D_1, \dots, D_{r-1}], \quad K' = [D'_0, D'_1, \dots, D'_{r-1}]$$

where for some integers $0 \leq p \leq q$:

- (1) $D_0 = D_1 = \dots = D_{p-1} = 0$,
- (2) $D'_0 = D'_1 = \dots = D'_{q-1} = 0$,
- (3) I, D_p, D'_q are linearly independent 3×3 matrices.

(The case $p = 0$ is to be interpreted as saying $D_0 \neq 0$. A similar statement applies for $q = 0$.)

Proof

Clearly we can achieve (1) and (2). We can also assume that these forms for our K, K' will not have $p = r$ or $q = r$, otherwise \mathcal{A} is a 2-generated commutative subalgebra and Gerstenhaber's theorem already applies. Suppose I and D_p are linearly dependent, say $D_p = aI$. Let $K'' = K - aW^p$. Then W, K'' , and K' generate the same subalgebra as W, K , and K' but now K'' has its first $p + 1$ blocks zero. Thus, we can assume I and D_p are independent. If I, D_p, D'_q are dependent, say $D'_q = aI + bD_p$, we can replace K' by $K'' = K' - aW^q - bKW^{q-p}$, which will have its first $q + 1$ blocks zero. Hence, we can assume that I, D_p, D'_q are independent. This establishes (3). \square

For a matrix $Z \in \mathcal{C}(W)$, we define the **W -translates** of Z to be the matrices

$$Z, ZW, ZW^2, \dots, ZW^{r-1}.$$

In top row notation, if $Z = [Z_0, Z_1, \dots, Z_{r-1}]$, then $ZW = [0, Z_0, Z_1, \dots, Z_{r-2}]$, $ZW^2 = [0, 0, Z_0, Z_1, \dots, Z_{r-3}]$, and so on. The blocks of Z are being translated to the right, hence our terminology. It is this translating effect of W under right multiplication which is fundamental to an understanding of our arguments.

Proposition 5.5.6

Under the assumptions in Lemma 5.5.5, the subalgebra $\mathcal{A} = F[W, K, K']$ is spanned (as a vector space) by I, K, K', K^2 and their W -translates.

Proof

Assume that K and K' have the form described in Lemma 5.5.5. Since W is 3-regular, the generalized Cayley–Hamilton Equation 5.2.1 gives

$$K^3 = A_0 + A_1K + A_2K^2$$

for some $A_0, A_1, A_2 \in F[W]$. Hence, K^3 and higher powers of K are in the span of I, K, K^2 and their W -translates. Similarly, $(K')^3$ and higher powers are in the span of $I, K', (K')^2$ and their W -translates. Therefore, to establish our proposition, it suffices to show that KK' and $(K')^2$ are in the span of I, K, K', K^2 and their W -translates. We will establish the result by induction¹⁶ on the sum $s = p + q$.

First suppose $s = 0$, that is, $p = q = 0$. The first leading edge subspace U_0 of \mathcal{A} is $U_0 = F[D_0, D'_0]$, which has dimension 3 because it is a commutative subalgebra of $M_3(F)$ containing the independent matrices I, D_0, D'_0 . (We have used Gerstenhaber's theorem here, note. See also Example 5.4.5.) By Theorem 5.4.4, U_0 is self-centralizing inside $M_3(F)$. Therefore by Proposition 3.4.4 (6), $\dim \mathcal{A} = 3r = n$. However, from the assumed forms of K and K' , we can see that the W -translates of I, K, K' are linearly independent and so these $3r = n$ matrices must form a basis for \mathcal{A} . So the result holds for $s = 0$. Next we assume $s > 0$, and consider two cases:

CASE I: $p < q$

Let X' be a 1-block pullback of K' and for the purposes of applying Proposition 5.5.3, view K as a 0-block pullback of itself (take $X = K$). Let $\pi : \mathcal{T} \rightarrow \mathcal{T}$ be the projection in Proposition 5.5.3 for $m = 0 + 1 = 1$. In the commutative corner algebra $F[\pi(W), \pi(K), \pi(X')]$, we see that $\pi(W)$ is still a 3-regular nilpotent Weyr matrix and the new sum “ s ” associated with $\pi(K), \pi(X')$ is smaller by 1 than our starting s . Also the (not necessarily strict) inequality is retained in the new “ p ” and “ q ,” and the three simplifications in Lemma 5.5.5 still hold. Hence, we can apply induction to $F[\pi(W), \pi(K), \pi(X')]$ and say it is spanned by $\pi(I), \pi(K), \pi(X'), \pi(K^2)$, and their $\pi(W)$ -translates. By Lemmas 5.5.2 and 5.5.4, we have (using l.c. as shorthand for linear combination):

$$\begin{aligned} K'K &= (X'K)W \\ &= \pi(X'K)W \\ &= [\pi(X')\pi(K)]W \end{aligned}$$

16. Here we use strong induction: (i) checking the result holds for $s = 0$, and (ii) that it holds for a given s if it holds for all smaller s . In the latter case, the given s may, of course, arise from various p and q , but that need not concern us. Alternatively, we could induct on p , combined with induction on q in the case $p = 0$. (Normally, “double induction” arguments should be avoided like the plague.)

$$\begin{aligned}
 &= [\text{l.c. of } \pi(I), \pi(K), \pi(X'), \pi(K^2), \text{ and their } \pi(W)\text{-translates}] W \\
 &= \text{l.c. of } W, KW, K', K^2W, \text{ and their } W\text{-translates.}
 \end{aligned}$$

A similar argument applies to $(K')^2$:

$$\begin{aligned}
 (K')^2 &= (X'W)^2 \\
 &= (\pi(X')W)^2 \\
 &= [\pi(X')]^2 W^2 \\
 &= [\text{l.c. of } \pi(I), \pi(K), \pi(X'), \pi(K^2), \text{ and their } \pi(W)\text{-translates}] W^2 \\
 &= \text{l.c. of } W^2, KW^2, K'W, K^2W^2, \text{ and their } W\text{-translates.}
 \end{aligned}$$

CASE II: $p = q$

This time we choose 1-block pullbacks X and X' for both K and K' and use the projection $\pi : \mathcal{T} \rightarrow \mathcal{T}$ in Proposition 5.5.3 for $m = 2$. The argument is now similar to Case I. We apply induction to the commutative algebra $F[\pi(W), \pi(X), \pi(X')]$ where now $\pi(W)$ is in Weyr form but is 2 blocks shorter in its Weyr structure, and the new “ p ” and “ q ” are smaller by 1 but the assumptions in Lemma 5.5.5 apply. Thus, by Lemmas 5.5.2 and 5.5.4 we have

$$\begin{aligned}
 K'K &= (X'X)W^2 \\
 &= \pi(X'X)W^2 \\
 &= [\pi(X')\pi(X)] W^2 \\
 &= [\text{l.c. of } \pi(I), \pi(X), \pi(X'), \pi(X^2), \text{ and their } \pi(W)\text{-translates}] W^2 \\
 &= \text{l.c. of } W^2, KW, K'W, K^2, \text{ and their } W\text{-translates.}
 \end{aligned}$$

$(K')^2$ is handled in a similar manner. This completes Case II and the inductive step. \square

Remark 5.5.7

The astute reader may have observed that one can avoid induction altogether in the proof of Proposition 5.5.6 by choosing a p -block pullback X of K , and a q -block pullback X' of K' . For let $m = s (= p + q)$. If $m \geq r$, then since $W^r = 0$ we have $KK' = XX'W^m = 0$ and $(K')^2 = (X')^2W^{2q} = 0$, whence the proposition holds in this case. Now suppose $m < r$ and let π be the projection as in Proposition 5.5.3. Observe that the commutative subalgebra $F[\pi(W), \pi(X), \pi(X')]$ is in case

“ $s = 0$,” whence is spanned by $\pi(I)$, $\pi(X)$, $\pi(X')$, and their $\pi(W)$ -translates. Therefore,

$$\begin{aligned}
 KK' &= XX'W^m \\
 &= \pi(XX')W^m \\
 &= [\pi(X)\pi(X')]W^m \\
 &= [\text{l.c. of } \pi(I), \pi(X), \pi(X'), \text{ and their } \pi(W)\text{-translates}]W^m \\
 &= \text{l.c. of } W^m, XW^m, X'W^m, \text{ and their } W\text{-translates} \\
 &= \text{l.c. of } W^m, KW^q, K'W^p, \text{ and their } W\text{-translates.}
 \end{aligned}$$

A similar argument works for $(K')^2$ after noting that $(K')^2 = (X')^2W^{2q} = (X')^2W^mW^{q-p}$. We are finished. Slicker though this argument is, we have a preference for the inductive proof, using just 1-block pullbacks, because this may give more insight into the construction of a minimal counterexample in other situations. (See the discussion prior to Example 3.5.3 in Chapter 3.) \square

We are now in a position to give our extension of Gerstenhaber’s theorem in the 3-regular case. To avoid any possible misquoting of our result, we spell out our hypotheses in full. The theorem is actually more general than it looks, because to establish the result (should it be true¹⁷) for commuting triples W, K, K' involving *any* 3-regular matrix W (not assumed nilpotent), it would suffice by Proposition 5.1.1 to consider just the case where W is nilpotent with a nonhomogeneous structure (such as $(3, 3, 3, 3, 2, 2)$). Although expressed in different terms, our theorem appears in the 2009 paper of Sethuraman and Šivic as their Corollary 4.7. Its proof is through algebraic geometry, using techniques similar to those we develop in Chapter 7. The key idea is that the “irreducibility of a certain algebraic variety” involving commuting triples of $n \times n$ matrices has vector space dimension implications for the subalgebra generated by such matrices. However, it seems unlikely that the reverse implication is true: for a given matrix order n , having the bound $\dim F[A_1, A_2, A_3] \leq n$ for all commuting $n \times n$ matrices A_1, A_2, A_3 may not necessitate irreducibility of the variety? Our techniques, on the other hand, offer some hope in both the nonhomogeneous 3-regular case and the homogeneous d -regular case when $d > 3$.

Theorem 5.5.8

Let $\mathcal{A} = F[W, K, K']$ be a 3-generated commutative subalgebra of $M_n(F)$ such that W is a 3-regular nilpotent matrix with a homogeneous Weyr structure. Then $\dim \mathcal{A} \leq n$.

17. The authors suspect that this is indeed true.

Proof

We return to our earlier notation and the simplifications in Lemma 5.5.5. We know $\dim \mathcal{A} = \dim U_0 + \dim U_1 + \cdots + \dim U_{r-1}$, the sum of the leading edge dimensions of \mathcal{A} (Theorem 3.4.3), and we need to show that on average $\dim U_i \leq 3$ because $n = 3r$. Let $k = \lfloor (r-1)/2 \rfloor$ (= largest integer less than or equal to $(r-1)/2$). By Proposition 3.4.4 (4), for $0 \leq i \leq k$, the matrices in U_i commute. Hence, since three is the maximum number of commuting linearly independent matrices in $M_3(F)$ (see Example 5.4.5), we must have

$$\dim U_i \leq 3 \quad \text{for } i = 0, 1, \dots, k.$$

Also, we know from Proposition 5.5.6 that \mathcal{A} is spanned by four matrices and their W -translates. It follows that $\dim U_i \leq 4$ for all i . (More generally, if \mathcal{A} is spanned by c matrices and their W -translates, then $\dim U_i \leq c$. Intuitively this is reasonable, although one needs to exercise care in a rigorous proof—a good exercise.)

If $\dim U_i \neq 3$ for all i , then we must have $\dim U_i \leq 2$ for $i = 0, 1, \dots, k$ and $\dim U_i \leq 4$ for $i = k+1, \dots, r-1$, so “on average” $\dim U_i \leq 3$ and we are finished. Therefore, we can assume some U_i has dimension 3. Let U_t be the first leading edge subspace for which $\dim U_t = 3$. By the same argument as in the previous paragraph, if $t \leq k$ then $\dim U_i = 3$ for $i = t, t+1, \dots, r-t-1$. This is because the U_i in this range centralize U_t but the latter generates a self-centralizing subalgebra of $M_3(F)$ (again consult Example 5.4.5). We now consider two cases.

Case 1: $2t < r$

As we did in Example 3.5.3, let us use the notation

d_0	d_1	d_2	\cdots	d_{r-1}
-------	-------	-------	----------	-----------

to indicate that the dimensions of the leading edge subspaces U_0, U_1, \dots, U_{r-1} are, respectively, d_0, d_1, \dots, d_{r-1} . Then in this case, the dimensions of the U_i are at worst

2	\cdots	2	3	\cdots	3	4	\cdots	4
\longleftarrow	t	\longrightarrow	\longleftarrow	$r-2t$	\longrightarrow	\longleftarrow	t	\longrightarrow

Thus,

$$\begin{aligned}
 \dim \mathcal{A} &= \sum_{i=0}^{t-1} \dim U_i + \sum_{i=t}^{r-t-1} \dim U_i + \sum_{i=r-t}^{r-1} \dim U_i \\
 &\leq 2t + 3(r-2t) + 4t \\
 &= 3r \\
 &= n.
 \end{aligned}$$

Case 2: $r \leq 2t$

In this case, we have

$$\begin{aligned}
 \dim \mathcal{A} &= \sum_{i=0}^{t-1} \dim U_i + \dim U_t + \sum_{i=t+1}^{r-1} \dim U_i \\
 &\leq 2t + 3 + 4(r - t - 1) \\
 &= (3r - 1) + (r - 2t) \\
 &\leq 3r - 1 \\
 &< n.
 \end{aligned}$$

This completes the proof of the theorem. □

When one assesses the performances of a good racehorse, or a good sports team, the key measure is not how many events they have won but who they have beaten. Against that standard, we judge the Weyr form to have acquitted itself very well in this opening event, the “Gerstenhaber Stakes.”

BIOGRAPHICAL NOTES ON CAYLEY AND HAMILTON

Arthur Cayley was born on August 16, 1821, in Richmond, England, as the son of a merchant. He was destined for the family business until his mathematics teacher at King’s College School persuaded his father to allow Arthur to follow his interest in mathematics. He studied at Trinity College, Cambridge, and before graduating first in his class in 1842, he had already published three papers in the *Cambridge Mathematical Journal*. After graduation, he taught at Cambridge for four years but, with no secured position there, he trained as a lawyer and was admitted to the bar in 1849. The legal profession gave him a comfortable living for the next 14 years but he regarded it only as a means of livelihood and continued to devote considerable time and energy to his prime interest, mathematics. Indeed, during these 14 years as an amateur he published almost 300 papers. In 1863 he took up the position of Sadlerian Professor of Pure Mathematics at Cambridge, with a dramatic decrease in income from his legal work. However, this move allowed him to concentrate entirely on mathematics, resulting in a lifetime total of 966 papers. These included the introduction of the concept of an abstract group, the foundation with his friend James Sylvester of the theory of invariants (for which they were dubbed “the invariant twins”), the beginning of matrix theory and determinants, and higher dimensional geometry. On November 19, 1857, Cayley wrote to Sylvester saying, “I have just obtained a theorem which appears to me *very remarkable*.” Now known as the Cayley–Hamilton theorem, this formally appeared in his

1858 tour de force *Memoir on the Theory of Matrices*. There he proved it for 2×2 matrices, stated that he had verified it for 3×3 s, but had “not thought it necessary to undertake the labour of a formal proof.” The first complete proof was given by Frobenius, 20 years later. Cayley died in Cambridge in 1895.

William Rowan Hamilton was born in Dublin, Ireland, on August 4, 1805, as the son of a Scottish lawyer. His father was often overseas but before he was five years old William had learned Latin, Greek, and Hebrew from his uncle James Hamilton. He became interested in mathematics when, aged 13, he met Zerah Colburn, an American known for his incredible ability in mental arithmetic. Hamilton's undergraduate years were spent at Trinity College, Dublin, and saw him receiving outstanding distinctions in science and classics. One of his finals examiners persuaded him to apply for the position of Royal Astronomer of Ireland. His application successful, he concurrently became Professor of Astronomy at Trinity College. However, he soon lost interest in astronomy and devoted his attentions to mathematics. As an undergraduate he had written a memoir named *Theory of Systems of Rays* in which he presented the characteristic function for optics. In the third supplement to this work he theoretically predicted conical refraction and gained great fame after this was soon verified experimentally. In 1833, at the Royal Irish Academy, he showed how the field of complex numbers could be expressed as “algebraic couples,” that is, ordered pairs of real numbers. He was knighted in 1835 and, after his algebraic couples insight, tried relentlessly for many years to extend his theory to triples. However, in 1843, while walking with his wife along the Royal Canal in Dublin, he realized that a fourth dimension was needed and he formulated the algebra of quaternions (a noncommutative field). His excitement in this led him to carve the quaternions' key equations $i^2 = j^2 = k^2 = ijk = -1$ with his penknife in the stone of the Canal's Broome Bridge. Hamilton's stake in the Cayley–Hamilton theorem arises in his *Lectures on Quaternions* where he shows that a rotation transformation in 3-dimensional space satisfies its own characteristic equation. He spent the rest of his life working on the quaternions and died on September 2, 1865.

Approximate Simultaneous Diagonalization

Simultaneous diagonalization of a finite collection of $n \times n$ matrices has long been recognized as a useful concept. And it continues to find new applications, such as to phylogenetic invariants for Markov models of sequence mutation.¹ Another recent and lovely application is to adaptive optics, that is, to methods that overcome the effects of distortion in imaging through a medium such as the atmosphere.² This is relevant to the functioning of earthbound astronomical telescopes³ (taking the “twinkling” out of the stars). In this chapter, we examine a second application of the Weyr form, to an approximate version of simultaneous diagonalization of complex matrices. This notion has also been used in some recent, dinkum applications.

In a 2003 study by Allman and Rhodes of phylogenetic invariants in biomathematics, the following question arose:⁴ Given A_1, A_2, \dots, A_k

1. See the Allman and Rhodes paper of 2003.

2. See the interesting 1998 article, and references therein, by Berman and Plemmons.

3. To some extent, this lessens the need for telescopes like the Hubble, that operate outside the earth’s atmosphere.

4. In the final published version of their work, Allman and Rhodes used another approach in terms of an equivalent condition of irreducibility of a certain complex affine variety of matrices. We discuss this connection fully in Chapter 7.

commuting $n \times n$ matrices over the complex numbers \mathbb{C} , can the matrices be perturbed by an arbitrarily small amount so that they become simultaneously diagonalizable? More specifically, given $\epsilon > 0$, are there $n \times n$ matrices E_i with $\|E_i\| < \epsilon$ and an invertible $n \times n$ matrix C such that $C^{-1}(A_i + E_i)C$ is diagonal for $i = 1, 2, \dots, k$? Any list of matrices with the property in question will be called **approximately simultaneously diagonalizable**, abbreviated to ASD. Notice that we do not assume commutativity for this definition. However, as we will see in Section 6.2, the ASD property actually implies the commutativity of the matrices A_i in the given list.

The ASD question was brought to our attention in 2003 by Mike Steel, an outstanding phylogeneticist at the University of Canterbury, New Zealand. Naïvely, some of us thought we could polish it off over morning tea. Later we discovered the question is directly related to some open questions in algebraic geometry. An attempt by O’Meara and Vinsonhaler in 2006 to attack the ASD question with just linear algebra led to their rediscovery of the Weyr form, which they termed the “H-form.” (That term has since been decommissioned.) So the origins of our book really lie with Mike Steel’s question. At Mike’s suggestion, Elizabeth Allman of the University of Alaska, Fairbanks, has kindly supplied us with a brief outline (for the nonexpert) of the ideas involved in phylogenetic invariants. We present this in Section 6.1. It is not essential reading, but we feel some readers will appreciate seeing another recent and topical “real-world application” of linear algebra, one for which the Weyr form has some relevance. Since the natural language for the study of phylogenetic invariants is algebraic geometry, Section 6.1 also serves as motivation for our Chapter 7.

Historically, the ASD property appears to have been studied only tangentially in the literature, mainly in connection with some problems in algebraic geometry. We take up this connection in Chapter 7. However, a more recent development by de Boer, Shekhtman, and others is to use the ASD property to study certain questions in multivariate interpolation. We won’t attempt to address that work in our book. In the present chapter, we study the ASD question by purely matrix-theoretic methods involving the Weyr form. The Weyr form provides a useful setting for constructing nice perturbations of commuting matrices. It is not our goal to give a definitive account of ASD, which is still being actively researched, but instead to illustrate the utility of the Weyr form in this type of analysis. Particularly useful are two properties that we established in Chapter 3: (1) the nice block upper triangular form of matrices that centralize a given nilpotent Weyr matrix; (2) the simultaneous triangularization property: given a finite list of commuting nilpotent matrices, we can put the first in Weyr form and make the rest strictly upper triangular, under a simultaneous similarity transformation.

As we demonstrate in Section 6.3, the ASD property fails in general for k commuting $n \times n$ matrices whenever $k \geq 4$ and $n \geq 4$. The earliest positive ASD result was established in 1955 by Motzkin and Taussky, who showed that any two commuting complex matrices (of any size) have the ASD property. We establish this in Section 6.8, after developing certain key tools in Sections 6.4, 6.5, and 6.7. It is when one has three commuting $n \times n$ complex matrices that the ASD question is still open. Currently the answer is known to be positive for $n \leq 8$, and negative when $n \geq 29$. We will establish the latter in Chapter 7 using the powerful techniques of algebraic geometry. In Section 6.12 of this chapter, we treat the cases $n \leq 5$, and make some comments on the recent work by Omladić, Han, and Šivic on the cases $n = 6, 7, 8$. The calculations involved in these latter cases can be very technical, and we will mostly avoid them.

There is another angle we can take with the ASD question. Presented with a fixed number k of commuting complex matrices, rather than ask for which size, $n \times n$, do *all* k commuting matrices possess the ASD property, we can ask when k *particular* $n \times n$ commuting matrices A_1, A_2, \dots, A_k have the ASD property. It turns out that this question is really one about the subalgebra $\mathcal{A} = \mathbb{C}[A_1, A_2, \dots, A_k]$ of $M_n(\mathbb{C})$ generated by the A_i and is independent of the generators: if A_1, A_2, \dots, A_k have ASD, then so too do all finite subsets of \mathcal{A} . An interesting consequence of this, covered in Section 6.3, is that the dimension of the subalgebra generated by $n \times n$ ASD matrices can't exceed n . As a corollary, using the Motzkin–Taussky theorem, we give in Section 6.8 a novel two-line proof of Gerstenhaber's theorem (studied in Chapter 5) over the complex field. Another purely algebraic necessary condition for complex $n \times n$ matrices A_1, A_2, \dots, A_k to satisfy ASD is that the centralizer of these matrices must have dimension at least n . This we cover in Section 6.6.

Continuing the theme of the previous paragraph, in Sections 6.9 and 6.10 we show that the ASD property holds for three commuting $n \times n$ matrices when one of them is 2-regular (that is, the eigenspaces of the matrix are at most 2-dimensional). This result is a corollary to a 1999 theorem by Neubauer and Sethuraman, who employed nontrivial methods of algebraic geometry. Our proof is purely matrix-theoretic, involving the Weyr form.

A classical problem, particular cases of which were studied in Chapter 5 over a general field, is that of finding an upper bound for the dimension over \mathbb{C} of $\mathbb{C}[A_1, A_2, \dots, A_k]$, the subalgebra (with identity) of the $n \times n$ complex matrices generated by commuting A_1, A_2, \dots, A_k . An old result of Schur says that the best upper bound in general is $\lfloor n^2/4 \rfloor + 1$. As noted above, n itself is a bound when the generators have ASD. In particular this holds when $k = 3$ and one of the A_i is 2-regular. In Section 6.11, we establish more generally that $\dim \mathbb{C}[A_1, A_2, \dots, A_k] \leq 5n/4$ for any k if one of the commuting A_i is 2-regular. This bound is sharp.

6.1 THE PHYLOGENETIC CONNECTION

Perhaps surprisingly, the questions of when a set of k commuting $n \times n$ complex matrices are simultaneously diagonalizable, or perturbable to a set of k simultaneously diagonalizable matrices, arise naturally in the setting of phylogenetic inference. Phylogenetics is the study of evolutionary relationships between a collection of organisms, typically species. In molecular phylogenetics, for example, one might strive to determine such historical relationships using aligned DNA sequences from a particular gene common to all the organisms of interest. The underlying belief is that sequence similarities and differences might hold the secret to understanding evolutionary relationships.

Though historically, evolutionary relationships amongst species might have been deduced from morphological data—beak size, beak shape, webbed feet/toes,—modern methods for DNA sequence analysis employ standard statistical methods that require a probabilistic model to describe the evolutionary descent of modern-day species from a common ancestor.

For example, consider the two rooted trees below depicting possible evolutionary relationships between three species of primates: human, chimpanzee, and gorilla. In both trees, we view time as progressing from the top (the past) to the bottom (the present), and bifurcations in the tree depict the origin of two new species from an ancestral one. The two trees in Figure 6.1 illustrate competing evolutionary hypotheses.

For DNA sequences, there are four nucleotides (or bases) ‘A’, ‘C’, ‘G’, and ‘T’. Now imagine that in aligned DNA sequences for these primates, we see the pattern ‘A A G’ at a particular site:

```

human:   .....A.....
chimp:   .....A.....
gorilla: .....G.....
  
```

The nucleotides occurring at this site in the extant species are assumed to have descended from a common ancestor. At the root of any tree relating the

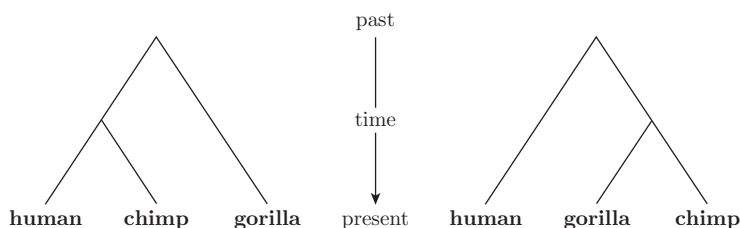


Figure 6.1

primates, then, the ancestral state must be an A, C, G, or T, and similarly at all of the other internal nodes of the tree, the extinct species these nodes represent must be in one of the four nucleotide states.

Suppose momentarily that the tree T on the left of Figure 6.1 is the “true” tree and the “true” ancestral state at the root was an ‘A’. (The use of the letter T to denote both the tree and the nucleotide shouldn’t confuse, and in any case, the former is in math font while the latter is in roman font.) Then the evolutionary history of this site might be depicted by either of the labeled trees shown in Figure 6.2. Of course, there are other possible labelings of internal nodes of the tree that give rise to the pattern ‘AAG’ that we observe at the leaves of the tree (the primates). And we can’t even be sure that the root was in state ‘A’.

Informally, we can introduce probabilities, or parameters, to explain our observation of the pattern ‘AAG’ in the aligned DNA of the primates. The idea is quite simple: we introduce a vector $\pi = (p_A \ p_C \ p_G \ p_T)$, the *root distribution* vector, that records the probability that each of the four nucleotides occurred at the root of the tree. For each edge in the tree, we need 16 conditional probabilities. For instance, for any directed edge $(u \rightarrow w)$ leading away from the root, we need a probability $p_{CA} = \mathbb{P}(w \text{ is in state A} \mid u \text{ is in state C})$. These 16 conditional probabilities describe the possible base changes that occur along the edge $(u \rightarrow w)$, conditioned on the initial state at u . With this collection of probabilities on T in hand, we can compute the expected value of the pattern ‘AAG’ at a site in the aligned sequences. (Simply sum over all possible labelings of internal nodes of the tree.) Usually, we call such an expected value a *pattern frequency* p_{AAG} and by introducing a probabilistic model to describe the evolutionary descent, we trust that for some choice of probabilities on some tree T , p_{AAG} is well approximated by the proportion of times we see the pattern ‘AAG’ in our data sequences. If so, then a statistical approach to molecular phylogenetic questions would find the tree T that ‘best’ approximates the observed pattern frequencies.

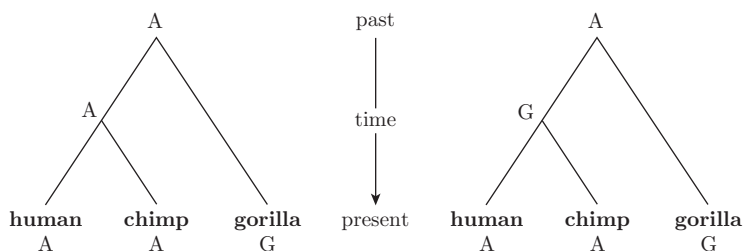


Figure 6.2

Of course, researchers are interested in relating a large number of species, even constructing a “tree of life” if possible, and modern techniques of phylogenetic inference are quite sophisticated, certainly more complicated than the toy example described here. For a more complete introduction to phylogenetics, the books *Phylogenetics* by Semple and Steel, and *Inferring Phylogenies* by Felsenstein provide thorough and well-written overviews of the field from both the mathematical and biological perspectives. See also the book *Mathematical Models in Biology* by Allman and Rhodes for more on the role of mathematics in biology.

For our purposes, making a connection to the simultaneous diagonalizability of a collection of matrices, we consider a much simpler situation, though it is still necessary to make formal our description of a probabilistic model on a tree. We build up to a description of a probabilistic model on an unrooted 3-leaf tree, relating species a , b , and c . Conveniently, the location of the root is immaterial to the computation of pattern frequencies, so we assume throughout that all trees are unrooted and we place the root in an ad hoc location to orient the edges of a tree. From the 3-leaf unrooted example, a motivated reader should be able to provide the correct definition for a probabilistic model on larger trees, such as those that might be required for data analysis.

Markov Models on Small Trees

We start with the simplest possible situation: a single-edge tree $e = (v \rightarrow a)$ relating a root v in the past to a present-day species a . The full set of parameters on this tree are (1) a root distribution vector $\pi = (p_A \ p_C \ p_G \ p_T)$, giving the distribution of nucleotides ‘A’, ‘C’, ‘G’, ‘T’ at the ancestral node v , and (2) a 4×4 Markov matrix $M_e = (m_{ij})$ whose entries are the conditional probabilities of various transitions occurring along the edge e . For instance, identifying ‘A’, ‘C’, ‘G’, ‘T’ with 1, 2, 3, 4, respectively, then $m_{23} = m_{CG} = \mathbb{P}(a \text{ is in state ‘G’} \mid v \text{ is in state ‘C’})$. For brevity, we usually write $m_{23} = \mathbb{P}(a = G \mid v = C)$, and let expressions like ‘ $a = G$ ’ denote assignments of states to nodes. Notice that all the transition probabilities associated with an edge have been grouped into a Markov transition matrix with rows summing to 1.

For a 3-leaf tree, like the one pictured in Figure 6.3, we proceed similarly. Assume that T is rooted at the internal node v . Then parameters for a model of sequence evolution on T are

- (1) a root distribution vector $\pi = (p_A \ p_C \ p_G \ p_T)$, giving the distribution of nucleotides at the root v , and
- (2) three 4×4 Markov matrices M_{va} , M_{vb} , and M_{vc} , one for each edge of the tree, giving the transition probabilities of various state changes along that edge of T .

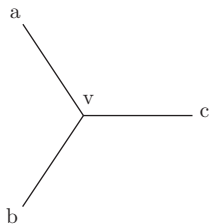


Figure 6.3 The 3-leaf tree T .

Now, with such parameters specified, we can compute pattern frequencies like p_{AAG} in terms of the entries of π and the entries of the three Markov matrices. To be concrete, let's compute p_{AAG} (or p_{113}). We find:

$$\begin{aligned} p_{\text{AAG}} &= p_{\text{A}} M_{va}(1, 1) M_{vb}(1, 1) M_{vc}(1, 3) + p_{\text{C}} M_{va}(2, 1) M_{vb}(2, 1) M_{vc}(2, 3) \\ &\quad + p_{\text{G}} M_{va}(3, 1) M_{vb}(3, 1) M_{vc}(3, 3) + p_{\text{T}} M_{va}(4, 1) M_{vb}(4, 1) M_{vc}(4, 3). \end{aligned}$$

To understand this pattern frequency, look more carefully at the second term in the sum. It corresponds to the (unknown) choice that the root v is in state 'C'. Using elementary rules of probability, we see that

$$\begin{aligned} p_{\text{C}} M_{va}(2, 1) M_{vb}(2, 1) M_{vc}(2, 3) &= p_{\text{C}} M_{va}(\text{C}, \text{A}) M_{vb}(\text{C}, \text{A}) M_{vc}(\text{C}, \text{G}) \\ &= p_{\text{C}} \mathbb{P}(a = \text{A} \mid v = \text{C}) \mathbb{P}(b = \text{A} \mid v = \text{C}) \mathbb{P}(c = \text{G} \mid v = \text{C}) \\ &= \mathbb{P}(v = \text{C}, a = \text{A}, b = \text{A}, c = \text{G}). \end{aligned}$$

Similarly, the three other summands contributing to p_{AAG} correspond to the possibility that each of the other three nucleotides (A, G, T) occurred at the root. (Remember the root is considered an ancestor of the current species a , b , and c and so its historic state is unknown.) Since we have summed over all possible assignments of nucleotides to the root, the pattern frequency p_{AAG} is exactly the joint probability $\mathbb{P}(a = \text{A}, b = \text{A}, c = \text{G})$ that pattern 'AAG' occurs at a site in DNA collected from the three species a , b , c depicted by the leaves of T .

For the above 3-leaf tree T and the Markov model on T , all $4^3 = 64$ pattern frequencies ($p_{\text{AAA}}, p_{\text{AAC}}, p_{\text{AAG}}, \dots, p_{\text{TTC}}, p_{\text{TTG}}, p_{\text{TTT}}$) can be computed similarly by summing over all possible assignments of states to the root v . Indeed, the pattern frequencies are *polynomial* formulas in the parameters of the model, and as we shall see in Chapter 7 this leads naturally to a well-known mathematical object known as a *parameterized algebraic variety*. Collecting all

pattern frequencies together, we obtain $P = \{p_{ijk}\}$, the **pattern frequency distribution** for patterns at the leaves of T .

Note that, since P is a probability distribution, by summing over all possible patterns we obtain a value of one:

$$p_{AAA} + p_{AAC} + \cdots + p_{TTG} + p_{TTT} - 1 = 0. \quad (*)$$

To researchers in phylogenetics, the left-hand side of $(*)$ is therefore an example of a **phylogenetic invariant** for our particular topological tree T in Figure 6.3—it is a polynomial f in 4^3 variables that vanishes when the expected pattern frequencies are substituted for the variables, *regardless of the values of the model parameters* π , $\{M_{va}, M_{vb}, M_{vc}\}$. Hence, if we label the variables by the pattern frequencies, then our phylogenetic invariant here is

$$f(p_{AAA}, \dots, p_{TTT}) = p_{AAA} + p_{AAC} + \cdots + p_{TTG} + p_{TTT} - 1.$$

Perhaps the form of this polynomial (and other such phylogenetic invariants) would be clearer if we had written $f(x_1, x_2, \dots, x_{64}) = x_1 + x_2 + \cdots + x_{64} - 1$, but then remembering where the substitutions go becomes a chore (e.g., remembering to substitute p_{ACG} in x_7). For emphasis we repeat: for *any* pattern frequency distribution P arising from model parameters on T ,

$$f(P) = 0.$$

A phylogenetic invariant is therefore associated with a particular tree T , but T will have infinitely many such invariants; they will form an ideal within the polynomial ring. (But importantly from the Hilbert Basis Theorem 7.3.1, this ideal is finitely generated.) And different trees will have some invariants in common.

Perhaps the polynomial invariant f appears not too interesting; it simply expresses the observation that P is a distribution of pattern frequencies. However, this viewpoint raises quite naturally a question of interest: *If P is a pattern frequency distribution arising from parameters on a tree T , what other polynomial relationships must hold between the pattern frequencies p_{pattern} ?*

We digress momentarily to shed some light on why invariants might be interesting in the context of phylogenetic inference. The ultimate goal, of course, would be to recover the phylogeny from the invariants. The main ideas date to the late 1980s in works of Cavender and Felsenstein, and Lake. From aligned sequence data collected from some number n of organisms, *observed* pattern frequencies, $\hat{p}_{A\dots A}$, etc., can easily be computed and collected into an *observed* distribution \hat{P} . Although we know that the sum of the observed

pattern frequencies is one (i.e., $f(\hat{P}) = 0$)), it is entirely unclear that there exists any choice of model parameters on an n -leaf tree T_n that might even approximately give rise to these observed frequencies. While a statistical method of phylogenetic inference of an evolutionary tree requires specifying a model of sequence evolution such as we have described here, there is absolutely no guarantee that sequence data is in accord with such a model.⁵ Though somewhat naïve on our part, perhaps invariants could in some way be used to assess fit of sequence data to model parameters. Indeed, if the observed distribution \hat{P} were well fit by model parameters on a fixed tree, then \hat{P} should be (close to) a zero of every phylogenetic invariant f for this model.

The ideas of Lake, and Cavender and Felsenstein, are even more elegant. They realized that if it were possible to find a phylogenetic invariant f_1 for a tree T_1 that was not a phylogenetic invariant for any other tree, and to find a phylogenetic invariant f_2 exclusive to T_2 , then the near vanishing of $f_1(\hat{P})$ might give evidence that topology T_1 is preferable to topology T_2 in explaining evolutionary relationships. Ramping this idea up a few notches, it might be possible to decide which topological relationships “best” describe sequence data using phylogenetic invariants.

As is often the case with novel and innovative ideas, the work of these researchers sparked quite a bit of interest, and many others have worked on finding phylogenetic invariants for models of sequence evolution on large n -leaf trees. This brings us closer to the connection with the simultaneous diagonalization of a collection of matrices.

Recalling that a phylogenetic invariant is a multivariate polynomial f that vanishes when evaluated at all pattern frequency distributions arising from model parameters, we return to the unrooted 3-leaf tree T discussed above. We next illustrate a technique of Allman and Rhodes (in their 2003 paper) for producing numerous phylogenetic invariants that involves the simultaneous diagonalization of matrices. The construction is quite technical and therefore optional; a reader may safely pass to the conclusion of this section if desired.

Phylogenetic Invariants from Simultaneous Diagonalizability of Three 4×4 Matrices

Given the joint distribution P of pattern frequencies on T of Figure 6.3, we view P not as a set of pattern frequencies but rather as a $4 \times 4 \times 4$ array. Equating the x -axis with species a , the y -axis with species b , and the z -axis with

5. Hence the well-known saying usually attributed to George Box stated here roughly, ‘All models are wrong, but some are useful.’

species c , we see that the pattern frequency $p_{\text{AAG}} = P(a = \text{A}, b = \text{A}, c = \text{G})$ is $P(1, 1, 3)$ in array notation.

Our interest is in looking at *slices* and sums of slices of the array P . For example, we consider the slice $P(\cdot, \cdot, 3)$, which can be viewed as the slice of P parallel to the xy -plane at a height $z = 3$. Indeed, $P(\cdot, \cdot, 3)$ is simply a 4×4 matrix where the row and column indices range over the four possible nucleotides for leaves a and b . Because $z = 3$, we are assuming that leaf c is fixed with nucleotide G appearing, while all possible combinations of nucleotides are possible at a and b . A moment's thought makes it clear that $P(\cdot, \cdot, 3)$ is the distribution of leaf patterns at a and b with $c = \text{G}$ so, for instance, the $(1, 1)$ entry is the probability of pattern 'AAG'. For brevity, we denote this 4×4 slice by $P_{ab\text{G}}$. The G in the third subscript indicates that leaf c is in state 'G', and the presence of the leaf names a and b indicates these indices are free to range over the four nucleotides. By analogy, we define three other slices $P_{ab\text{A}}$, $P_{ab\text{C}}$, and $P_{ab\text{T}}$ which are 4×4 matrices parallel to $P_{ab\text{G}}$ and correspond to particular assignments of a nucleotide to leaf c .

We are also interested in sums of these slices and define

$$P_{ab\bullet} = P_{ab\text{A}} + P_{ab\text{C}} + P_{ab\text{G}} + P_{ab\text{T}},$$

a 4×4 matrix of joint probabilities for leaves a and b . For those knowing probability theory, this is the marginalization of P over c , as it gives the distribution of a and b for any nucleotide appearing at leaf c . Assume now that $\det(P_{ab\bullet}) \neq 0$, so that $P_{ab\bullet}$ is an invertible matrix, and that all the entries of π are positive.

The key to our construction of phylogenetic invariants is a remarkable set of equations that express certain products of these slices and sums of slices of P in terms of the parameters π , M_{va} , M_{vb} , M_{vc} . For instance,

$$(P_{ab\bullet})^{-1}P_{ab\text{A}} = \left[M_{vb}^{-1}(\text{diag}(\pi))^{-1}(M_{va}^T)^{-1} \right] \left[M_{va}^T \text{diag}(\pi) \Lambda_{c,1} M_{vb} \right],$$

where $\Lambda_{c,1}$ is a diagonal matrix with diagonal entries given by the first column of M_{vc} , and the superscript 'T' denotes matrix transpose. This simplifies to

$$(P_{ab\bullet})^{-1}P_{ab\text{A}} = M_{vb}^{-1} \Lambda_{c,1} M_{vb}, \quad (1)$$

a diagonalization of the matrix product $(P_{ab\bullet})^{-1}P_{ab\text{A}}$.

Letting the nucleotide at leaf c be each of the three other possible nucleotides, we find three analogous equations for diagonalizations of matrix products:

$$\begin{aligned}(P_{ab\bullet})^{-1}P_{abC} &= M_{vb}^{-1}\Lambda_{c,2}M_{vb}, \\ (P_{ab\bullet})^{-1}P_{abG} &= M_{vb}^{-1}\Lambda_{c,3}M_{vb}, \\ (P_{ab\bullet})^{-1}P_{abT} &= M_{vb}^{-1}\Lambda_{c,4}M_{vb}.\end{aligned}\tag{2}$$

We are now ready for the main point: the matrix products $(P_{ab\bullet})^{-1}P_{abA}$, $(P_{ab\bullet})^{-1}P_{abC}$, $(P_{ab\bullet})^{-1}P_{abG}$, and $(P_{ab\bullet})^{-1}P_{abT}$ are **simultaneously diagonalizable** by M_{vb}^{-1} ! A first consequence of this is that these matrices all commute pairwise (see Proposition 6.2.6 to come):

$$\left[(P_{ab\bullet})^{-1}P_{abA}\right]\left[(P_{ab\bullet})^{-1}P_{abC}\right] = \left[(P_{ab\bullet})^{-1}P_{abC}\right]\left[(P_{ab\bullet})^{-1}P_{abA}\right], \text{ etc.}$$

Moreover, after a little algebra, we can manipulate this set of equations to find

$$\left[P_{abA}\right]\left[(\det(P_{ab\bullet})(P_{ab\bullet})^{-1})P_{abC}\right] = \left[P_{abC}\right]\left[(\det(P_{ab\bullet})(P_{ab\bullet})^{-1})P_{abA}\right] \tag{3}$$

and five other similar equations. A pleasing feature of Equation (3) is that each matrix entry is polynomial in the pattern frequencies P and, thus, by taking the difference of each of the 16 corresponding matrix entries, we have found 16 phylogenetic invariants, which a joint distribution on the tree T must satisfy. These invariants are polynomials of degree at most 5 (in 64 variables). All this from the observation that $(P_{ab\bullet})^{-1}P_{abX}$ for $X \in \{A, C, G, T\}$ are simultaneously diagonalizable.

In truth, we need only consider the three Equations (2), since Equation (1) follows from these as $P_{abA} = P_{ab\bullet} - (P_{abC} + P_{abG} + P_{abT})$. We have arrived at the conclusion that the question of the simultaneous diagonalizability of three 4×4 matrices is intimately related to the existence of polynomial relationships on a phylogenetic distribution P on T .

A Phylogenetic Connection with ASD as well

Looking still further ahead, we comment that the existence of phylogenetic invariants for phylogenetic models on trees—polynomials that vanish at pattern frequencies—also connects up with the notion of approximate simultaneous diagonalization (ASD) of sets of three 4×4 matrices. The details of this would

take us beyond our stated brief and do require a more thorough understanding of algebraic varieties, such as that developed in Chapter 7. In essence, ASD is used to verify that certain sets of phylogenetic invariants for a tree T are “strong” (have as much distinguishing power as the set of all invariants for T). The key properties used are the following:

- (1) Suppose that for a given k and n , all k -tuples of commuting complex $n \times n$ matrices are ASD. Let f be a polynomial function on the entries of k -tuples (A_1, A_2, \dots, A_k) of $n \times n$ matrices. If f vanishes on k -tuples of simultaneously diagonalizable matrices, then f must vanish on all k -tuples of commuting matrices (and conversely).
- (2) All triples of commuting 4×4 matrices are ASD (see Section 6.12).

The interested reader can refer to Section 6 of the Allman–Rhodes 2003 paper for the full details.

6.2 BASIC RESULTS ON ASD MATRICES

The concept of perturbing the entries of a complex matrix is an old one, belonging to the general area of mathematics called Perturbation Theory. Our use of the term “perturbation” implicitly means the change can be arbitrarily small (often spelled out as an ϵ -perturbation). Our interest lies in perturbing $n \times n$ complex matrices so that they become simultaneously diagonalizable. This section presents some appetizers. The real meat comes later.

Throughout this chapter, our algebraically closed ground field is the field \mathbb{C} of complex numbers. Most of what we do, however, could be formulated over any algebraically closed field of characteristic zero. We assume the reader is comfortable with the standard norm $\| \cdot \|$ on complex m -space \mathbb{C}^m :

$$\| \mathbf{a} \| = \| (a_1, a_2, \dots, a_m) \| = \sqrt{|a_1|^2 + |a_2|^2 + \dots + |a_m|^2}$$

for all $\mathbf{a} = (a_1, a_2, \dots, a_m) \in \mathbb{C}^m$. There is, of course, associated with this norm the standard Euclidean metric given by $d(\mathbf{a}, \mathbf{b}) = \| \mathbf{a} - \mathbf{b} \|$ for all $\mathbf{a}, \mathbf{b} \in \mathbb{C}^m$. Our “measure of closeness” will always be relative to this metric. So in expressions such as “an ϵ -perturbation of \mathbf{a} ” we mean a change of \mathbf{a} to some \mathbf{b} for which $\| \mathbf{a} - \mathbf{b} \| < \epsilon$ (or less than a constant multiple of ϵ). We can apply this notion of a norm (and associated metric) to $n \times n$ complex matrices simply by treating $A \in M_n(\mathbb{C})$ as a vector in \mathbb{C}^{n^2} in the standard way. (Thus, $\|A\|$ is the square root of the sum of the absolute values squared of the matrix entries.) The key

property we use of the norm in this matrix setting⁶ (apart from the triangle inequality $\|A + B\| \leq \|A\| + \|B\|$) is the submultiplicative inequality

$$\|AB\| \leq \|A\| \|B\|$$

for all $n \times n$ matrices A and B .

Let us now be quite clear about what is meant by *approximately simultaneously diagonalizable* matrices. Recall that a single matrix $B \in M_n(\mathbb{C})$ is **diagonalizable** if B is similar to a diagonal matrix. A collection B_1, B_2, \dots, B_k of $n \times n$ matrices is said to be **simultaneously diagonalizable** if the matrices are not only individually diagonalizable but in fact become diagonal under some common similarity transformation: for some invertible matrix C , we have $C^{-1}B_iC$ is diagonal for each $i = 1, 2, \dots, k$.

Definition 6.2.1: Complex $n \times n$ matrices A_1, A_2, \dots, A_k are said to be **approximately simultaneously diagonalizable** (abbreviated ASD) if, for each positive real number ϵ , there exist complex $n \times n$ matrices B_1, B_2, \dots, B_k that are simultaneously diagonalizable and satisfy

$$\|B_i - A_i\| < \epsilon \text{ for all } i = 1, 2, \dots, k. \quad \square$$

Example 6.2.2

A single matrix $A \in M_n(\mathbb{C})$ is always approximately diagonalizable. Put another way, A is a limit (relative to the matrix norm $\|\cdot\|$) of diagonalizable matrices. To see this, choose an invertible matrix P such that $P^{-1}AP$ is an upper triangular matrix, say, $T = (t_{ij})$. Given an $\epsilon > 0$, we can ϵ -perturb the diagonal entries $t_{11}, t_{22}, \dots, t_{nn}$ of T so that they become distinct. Now the perturbed matrix \bar{T} has n distinct eigenvalues and therefore \bar{T} is diagonalizable. Let $B = P\bar{T}P^{-1}$, which is clearly also diagonalizable. Finally, observe that B is a perturbation of A , though not for the same epsilon. But by the usual “epsilon-delta” argument of analysis we can arrange for B to be an ϵ -perturbation of A by starting with some smaller ϵ' . A more elegant way of viewing this is to observe that the conjugation map

$$\theta : M_n(\mathbb{C}) \longrightarrow M_n(\mathbb{C}), \quad X \longmapsto PXP^{-1}$$

is continuous in the usual (metric) topology. Therefore, for all sufficiently small ϵ' -perturbations \bar{T} of T , continuity of θ ensures that $\theta(\bar{T}) = B$ is an ϵ -perturbation

6. Some readers may prefer to use another equivalent norm, such as the operator norm, or simply the maximum of the absolute values of the matrix entries. In the latter case, for fixed n , we still have $\|AB\| \leq c\|A\| \|B\|$ for some constant c , which is fine for our arguments.

of $\theta(T) = A$ for the given ϵ . In future, we will tend to slide over this point, but let us spell it out here. If we let $\epsilon' = \epsilon / \|P\| \|P^{-1}\|$, then when $\|\bar{T} - T\| < \epsilon'$ we have

$$\begin{aligned} \|B - A\| &= \|P\bar{T}P^{-1} - PTP^{-1}\| \\ &= \|P(\bar{T} - T)P^{-1}\| \\ &\leq \|P\| \cdot \|\bar{T} - T\| \cdot \|P^{-1}\| \\ &< \|P\| \cdot \|P^{-1}\| \cdot \epsilon' \\ &= \epsilon. \end{aligned} \quad \square$$

Example 6.2.3

Fix $A \in M_n(\mathbb{C})$. Then any finite collection A_1, A_2, \dots, A_k of matrices in the subalgebra $\mathbb{C}[A]$ of $M_n(\mathbb{C})$ generated by A has the ASD property. To see this, note first that each A_i is some polynomial in A , say $A_i = p_i(A)$, and we know from Example 6.2.2 that A can be perturbed to a diagonalizable matrix \bar{A} . By the same sort of continuity argument as in Example 6.2.2, the matrices $B_i = p_i(\bar{A})$ are perturbations of the A_i for $i = 1, 2, \dots, k$. The B_i are also simultaneously diagonalizable because they are all polynomials in the fixed diagonalizable matrix \bar{A} . Thus, A_1, A_2, \dots, A_k are ASD. \square

Example 6.2.4

Consider the following 3×3 matrices:

$$A_1 = \begin{bmatrix} 3 & 0 & 1 \\ 1 & 5 & 4 \\ -1 & -3 & -2 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -1 & 0 & -1 \\ -1 & -3 & -4 \\ 1 & 3 & 4 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 5 & -3 & 0 \\ 3 & 8 & 9 \\ -3 & -6 & -7 \end{bmatrix}.$$

The second and third matrices are polynomials in the first: more precisely, $A_2 = 2I - A_1$ and $A_3 = -A_1 + A_1^2$. Hence, by Example 6.2.3, we know that A_1, A_2, A_3 are ASD. Let's rework the argument to find explicit perturbations to simultaneously diagonalizable matrices B_1, B_2, B_3 . Conjugating A_1 by

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix}$$

produces the upper triangular matrix

$$T = P^{-1}A_1P = \begin{bmatrix} 2 & -1 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}.$$

We can perturb T to a diagonalizable matrix by taking

$$\overline{T} = \begin{bmatrix} 2 + \epsilon & -1 & 1 \\ 0 & 2 - \epsilon & 3 \\ 0 & 0 & 2 \end{bmatrix}.$$

(\overline{T} has eigenvalues $2 + \epsilon$, $2 - \epsilon$, 2 , which are distinct for $\epsilon > 0$.) We can now take

$$B_1 = P\overline{T}P^{-1} = \begin{bmatrix} 3 + \epsilon & 0 & 1 \\ 1 + 2\epsilon & 5 - \epsilon & 4 \\ -1 - 2\epsilon & -3 + \epsilon & -2 \end{bmatrix},$$

$$B_2 = 2I - B_1 = \begin{bmatrix} -1 - \epsilon & 0 & -1 \\ -1 - 2\epsilon & -3 + \epsilon & -4 \\ 1 + 2\epsilon & 3 - \epsilon & 4 \end{bmatrix},$$

$$B_3 = -B_1 + B_1^2 = \begin{bmatrix} 5 + 3\epsilon + \epsilon^2 & -3 + \epsilon & \epsilon \\ 3 + 6\epsilon & 8 - 5\epsilon + \epsilon^2 & 9 - 2\epsilon \\ -3 - 6\epsilon & -6 + 5\epsilon - \epsilon^2 & -7 + 2\epsilon \end{bmatrix},$$

and these are perturbations of our original A_1, A_2, A_3 to simultaneously diagonalizable matrices. We could calculate an explicit invertible matrix $C = C(\epsilon)$, whose entries are rational functions of ϵ and which simultaneously conjugates B_1, B_2, B_3 to diagonal matrices. Such C are in general not pretty.

Notice that none of our A_i in this example is actually diagonalizable, because each has only a single eigenvalue but is not a scalar matrix. Why must it follow that

$$\lim_{\epsilon \rightarrow 0} C(\epsilon)$$

either doesn't exist or is not invertible? □

Recall that a *1-regular* (or *nonderogatory*) matrix A is one whose eigenspaces are 1-dimensional. (This doesn't mean its eigenvalues are distinct, although that would certainly be sufficient.) For such A , we know that matrices that commute with A must be polynomials in A . See Chapter 3, Proposition 3.2.4.

Proposition 6.2.5

Complex $n \times n$ matrices A_1, A_2, \dots, A_k are ASD if and only if they can be perturbed to commuting matrices X_1, X_2, \dots, X_k , one of which is 1-regular.

Proof

If ASD holds, then the A_i can be perturbed to simultaneously diagonalizable B_i . Suppose $C \in GL_n(\mathbb{C})$ conjugates B_i to diagonal D_i for $i = 1, 2, \dots, k$, that is, $C^{-1}B_iC = D_i$. Perturb the diagonal entries of D_1 to make them distinct and call the new matrix \overline{D}_1 . Then \overline{D}_1 is 1-regular and still commutes with the other D_i . Now $C\overline{D}_1C^{-1}, B_2, \dots, B_k$ (the result of applying the inverse conjugation to $\overline{D}_1, D_2, \dots, D_k$) are commuting perturbations of A_1, A_2, \dots, A_k with the first matrix 1-regular. Note that by perturbing the diagonal entries of the other D_i , we could make all the perturbed matrices commuting and 1-regular.

Conversely, suppose A_1, A_2, \dots, A_k can be perturbed to commuting X_1, X_2, \dots, X_k with say X_1 a 1-regular matrix. By 1-regularity, each X_i is a polynomial in X_1 . Therefore, by Example 6.2.3, X_1, X_2, \dots, X_k are ASD, hence so too are A_1, A_2, \dots, A_k . \square

The following proposition is well known and important in both directions.⁷

Proposition 6.2.6

Over any field F , matrices $B_1, B_2, \dots, B_k \in M_n(F)$ are simultaneously diagonalizable if and only if they are individually diagonalizable and commute.

Proof

Suppose the matrices are simultaneously diagonalizable with $C^{-1}B_iC$ diagonal for some invertible C and for $i = 1, 2, \dots, k$. Since diagonal matrices commute, we see that the $C^{-1}B_iC$ commute. Hence, so do the $B_i = C(C^{-1}B_iC)C^{-1}$, because conjugation is an algebra automorphism.

Conversely, suppose the B_i are diagonalizable and commute. If none of the B_i has at least two distinct eigenvalues, then they are all scalar matrices and so trivially are simultaneously diagonalizable. (To see this, note that (i) similar matrices have the same eigenvalues, (ii) the eigenvalues of a diagonal matrix are its diagonal entries, (iii) a diagonal matrix with no two distinct eigenvalues is a scalar matrix, and (iv) the only matrix similar to a scalar matrix is the same scalar matrix.) On the other hand, if some B_j has at least two distinct eigenvalues, then there is a simultaneous similarity transformation under which all the B_i are (nontrivially) block diagonal with matching block sizes. (See Proposition 5.1.1.) The matching blocks must commute and are also diagonalizable (because their minimal polynomial divides that of the parents, hence is a product of distinct linear factors). By induction on n , the matching blocks are simultaneously diagonalizable, whence so are the B_i . \square

7. The result was certainly known to McCoy in the 1930s (and also stated in the case of $k = 2$ by Cherubino in 1936). But it probably dates back much further, possibly as far back as Frobenius in the 1890s.

Remark 6.2.7

Another interesting characterization⁸ of the simultaneous diagonalizability of B_1, B_2, \dots, B_k is that the B_i are each polynomials $p_i(B)$ in some common diagonalizable matrix B . (This contrasts sharply with Example 5.1.4.) Of course, it suffices to establish this when the B_i are actually diagonal. In this case, let $B = \text{diag}(b_1, b_2, \dots, b_n)$ be any diagonal matrix with distinct diagonal entries. Then any other diagonal matrix $D = \text{diag}(d_1, d_2, \dots, d_n)$ is polynomial in B , namely $D = p(B)$ where $p(x) \in F[x]$ satisfies

$$p(b_i) = d_i \text{ for } i = 1, \dots, n$$

(such as given by the Lagrange interpolation formula).⁹ □

A natural question is whether the “approximate” version of Proposition 6.2.6 holds. By Example 6.2.2, every matrix is approximately diagonalizable, so the question is whether ASD is equivalent to some sort of “approximate commutativity.” The answer is no because we will see examples in Section 6.3 of commuting matrices that fail ASD. However, (full) commutativity¹⁰ is necessary for ASD:

Proposition 6.2.8

If A_1, A_2, \dots, A_k are ASD, then $A_i A_j = A_j A_i$ for all i, j .

Proof

Suppose, for example, that A_1 and A_2 do not commute. Since the commutator mapping $(X, Y) \mapsto [X, Y] = XY - YX$ from $M_n(\mathbb{C}) \times M_n(\mathbb{C})$ to $M_n(\mathbb{C})$ is continuous in the standard topology, and $[A_1, A_2] \neq 0$, there exists $\epsilon > 0$ such that $[B_1, B_2] \neq 0$ for all B_1, B_2 with $\|B_i - A_i\| < \epsilon$ for $i = 1, 2$. For such a pair B_1 and B_2 , the nonzero commutator says they do not commute so they cannot be simultaneously diagonalizable by Proposition 6.2.6. This contradicts the ASD hypothesis, completing the proof. □

On the one hand, Proposition 6.2.8 is good news. It provides a simple, purely algebraic, necessary condition for ASD. We will see two other important such conditions in Sections 6.3 and 6.6. On the other hand, the result is also sobering, for the following reason. If we start off with commuting matrices A_1, A_2, \dots, A_k , it is very difficult in general to make nontrivial perturbations of these matrices to *commuting* matrices (and, recall

8. This appears in the 1951 paper of Drazin, Dungey, and Gruenberg.

9. See, for example, p. 124 of Hoffman and Kunze's *Linear Algebra, 2nd Edition*.

10. Since commutativity is a topologically closed condition, “approximate commutativity” really has no interpretation other than full commutativity.

by Proposition 6.2.6, simultaneously diagonalizable matrices must commute). The reason is topological. If two matrices B_1 and B_2 don't commute, then there is a neighborhood of each such that no perturbations of B_1 and B_2 within these neighborhoods will commute (failure of commutativity is an open condition). So it is not good enough to start off with some perturbations of commuting matrices that "almost commute," in the hope of fudging it in another step. One has to get it right in one go. For instance,

$$B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } B_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

don't commute, and one can check (by looking at the $(1, 2)$ entry of the commutator $[B_1, B_2]$) that any perturbations of them

$$\bar{B}_1 = \begin{bmatrix} 1 + \epsilon_1 & \epsilon_2 \\ \epsilon_3 & \epsilon_4 \end{bmatrix} \text{ and } \bar{B}_2 = \begin{bmatrix} \epsilon_5 & 1 + \epsilon_6 \\ \epsilon_7 & \epsilon_8 \end{bmatrix}$$

won't commute either when (crudely) $\|\bar{B}_i - B_i\| < 0.1$ for $i = 1, 2$.

6.3 THE SUBALGEBRA GENERATED BY ASD MATRICES

Our second, and deeper, necessary condition for ASD of a collection of $n \times n$ matrices is that the subalgebra (with identity) of $M_n(\mathbb{C})$ that these matrices generate can have dimension at most n . Notice that, like the earlier necessary condition of commutativity, this second condition is also a purely algebraic condition. In some ways, the result can be viewed as an extension of Gerstenhaber's theorem in the complex case. (This will become clearer after the Motzkin–Tausky theorem in Section 6.8.) We proceed through some preliminary results to establish this property. The first says that the ASD condition for k given $n \times n$ matrices is really a statement about the subalgebra they generate, independent of the generators of this subalgebra. This is often a useful point of view, and it generalizes Example 6.2.3.

Proposition 6.3.1

Let \mathcal{A} be a commutative subalgebra (with identity) of $M_n(\mathbb{C})$ and suppose A_1, \dots, A_k generate \mathcal{A} as an algebra. If A_1, \dots, A_k are ASD, then so also is any finite set of matrices in \mathcal{A} .

Proof

There exists a \mathbb{C} -vector space basis for \mathcal{A} of monomials M_1, \dots, M_r in the A_i , say, of degree at most d . By taking $M_1 = I$, we can assume M_1 has degree 0 and the other M_j have positive degree. Given $\epsilon > 0$, let $b = \max\{\|A_1\|, \dots, \|A_k\|, 1\}$ and

$\epsilon' = \epsilon/2^d b^{d-1}$. Suppose, using the ASD hypothesis, that $A_1 + E_1, \dots, A_k + E_k$ are simultaneously diagonalizable approximations of A_1, A_2, \dots, A_k with $\|E_i\| < \epsilon'$ for $i = 1, 2, \dots, k$. Substitute $A_i + E_i$ for A_i in the monomials M_j to obtain monomials M'_j in the $A_i + E_i$. We can expand M'_j as a sum of the monomial M_j and monomial terms involving error terms E_i as well as the original matrices A_i . Each of the error term monomials involves at most $d - 1$ matrices A_i and there are $2^d - 1$ such terms. Thus, $\|M'_j - M_j\| < 2^d b^{d-1} \epsilon' = \epsilon$. Note that the matrices M'_j are simultaneously diagonalizable since the matrices $A_i + E_i$ are. Therefore, the basis M_1, \dots, M_r can be approximated by simultaneously diagonalizable matrices.

Now let $\{X_1, \dots, X_s\}$ be a finite subset of \mathcal{A} . For $i = 1, \dots, s$ write $X_i = \sum_{j=1}^r c_{ij} M_j$ and let $c = \max_{i,j} \{|c_{ij}|\}$. Given $\epsilon > 0$, let M'_1, \dots, M'_r be simultaneously diagonalizable ϵ -approximations of M_1, \dots, M_r . Set $X'_i = \sum_{j=1}^r c_{ij} M'_j$. Then X'_1, \dots, X'_s are simultaneously diagonalizable and

$$\|X'_i - X_i\| = \left\| \sum_{j=1}^r c_{ij} (M'_j - M_j) \right\| \leq \sum_{j=1}^r |c_{ij}| \|M'_j - M_j\| < rc\epsilon.$$

Hence, X_1, \dots, X_s can be approximated by simultaneously diagonalizable matrices, as asserted. \square

Lemma 6.3.2

If A_1, A_2, \dots, A_k are linearly independent in $M_n(\mathbb{C})$, then there exists $\epsilon > 0$ such that if B_i satisfies $\|B_i - A_i\| < \epsilon$ for $i = 1, 2, \dots, k$, then B_1, B_2, \dots, B_k are also linearly independent.

Proof

We can view the matrices as vectors in \mathbb{C}^{n^2} . So it is enough to establish the result for k independent vectors in general m -space \mathbb{C}^m , in fact for m linearly independent vectors v_1, \dots, v_m in \mathbb{C}^m (after expanding the original set to a basis). Let M be the $m \times m$ matrix with v_1, \dots, v_m as its columns. Since the determinant function $\det : M_m(\mathbb{C}) \rightarrow \mathbb{C}$ is continuous (in the topology induced by the norm), and $\det M \neq 0$, there is an open neighborhood \mathcal{N} of M such that $\det X \neq 0$ for all $X \in \mathcal{N}$. Since any such X has independent columns, the result follows. \square

Now we come to the main act of this part of the show.

Theorem 6.3.3

If a commutative subalgebra \mathcal{A} of $M_n(\mathbb{C})$ has a finite set of generators that can be approximated by simultaneously diagonalizable matrices, then $\dim \mathcal{A} \leq n$.

Proof

Let $r = \dim \mathcal{A}$ and let $\{B_1, \dots, B_r\}$ be a vector space basis for \mathcal{A} . By Proposition 6.3.1, B_1, \dots, B_r can be approximated by simultaneously diagonalizable matrices B'_1, \dots, B'_r . Moreover, by Lemma 6.3.2, we can arrange for B'_1, \dots, B'_r to be linearly independent. Let C be an invertible matrix such that, for each i , $C^{-1}B'_iC = D_i$, a diagonal matrix. Since D_1, \dots, D_r are linearly independent members of the n -dimensional space of diagonal matrices, $r = \dim \mathcal{A} \leq n$. \square

An attractive question is whether the converse of Theorem 6.3.3 holds: if A_1, A_2, \dots, A_k are commuting $n \times n$ matrices that generate a subalgebra of dimension at most n , must the matrices be ASD? In 2008, de Boer and Shekhtman provided a counterexample by constructing 16 commuting complex 17×17 matrices that generate a subalgebra of dimension 17 but are not ASD. Boris Shekhtman has also informed the authors that for large n , one can demonstrate the *existence* of three commuting $n \times n$ matrices A_1, A_2, A_3 such that $\dim \mathbb{C}[A_1, A_2, A_3] = n$ but A_1, A_2, A_3 fail the ASD property.¹¹

With the aid of Theorem 6.3.3, we can now show that, in general, ASD fails for k commuting $n \times n$ matrices whenever $k \geq 4$ and $n \geq 4$. Clearly, it suffices to consider the case $k = 4$ and $n \geq 4$.

Example 6.3.4

For each $n \geq 4$, there exist four commuting $n \times n$ complex matrices A_1, A_2, A_3, A_4 that fail the ASD property.

First, we consider the case $n = 4$, and let $E_1 = e_{13}$, $E_2 = e_{14}$, $E_3 = e_{23}$, $E_4 = e_{24}$. (Here e_{ij} denotes the matrix unit with a 1 in the (i, j) position and zeros elsewhere.) Notice that all the products $E_i E_j$ are zero, whence E_1, \dots, E_4 generate the following commutative subalgebra (with identity):

$$\begin{aligned} \mathcal{A} &= \text{set of scalar matrices} + \text{linear span of } E_1, \dots, E_4 \\ &= \left\{ \begin{bmatrix} a & 0 & b & c \\ 0 & a & d & e \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & a \end{bmatrix} : a, b, c, d, e \in \mathbb{C} \right\}. \end{aligned}$$

Since $\dim \mathcal{A} = 5 > 4 = n$, by Theorem 6.3.3, E_1, \dots, E_4 fail the ASD property.

11. In a private communication, Boris Shekhtman says he would love to see explicit examples of such triples, especially for small n . His interest in this relates to multivariate interpolation.

Now suppose $n > 4$. Let $m = n - 4$. We construct the A_i as block diagonal matrices $\text{diag}(E_i, F_i)$ where the E_i are the 4×4 matrices above and the F_i are suitable $m \times m$ matrices. Namely, we take F_1 to be an invertible $m \times m$ matrix whose group order in $GL_m(\mathbb{C})$ is m and whose first m powers are independent (such as the permutation matrix corresponding to the cycle $(1\ 2\ \dots\ m)$); and the other F_i we take to be the zero matrix. Clearly, the A_i commute because the E_i commute and the F_i commute. Notice that the subalgebra $\mathcal{A} = \mathbb{C}[A_1, A_2, A_3, A_4]$ contains the idempotent $A_1^{2m} = \text{diag}(0, I_m)$ (and also its complement $\text{diag}(I_4, I_m) - \text{diag}(0, I_m) = \text{diag}(I_4, 0)$ because our subalgebras always contain the identity). Therefore, we have a direct product decomposition of algebras

$$\mathcal{A} \cong \mathbb{C}[E_1, E_2, E_3, E_4] \times \mathbb{C}[F_1, F_2, F_3, F_4],$$

which implies

$$\begin{aligned} \dim \mathcal{A} &= \dim (\mathbb{C}[E_1, E_2, E_3, E_4] \times \mathbb{C}[F_1, F_2, F_3, F_4]) \\ &= \dim \mathbb{C}[E_1, E_2, E_3, E_4] + \dim \mathbb{C}[F_1, F_2, F_3, F_4] \\ &= 5 + m > n. \end{aligned}$$

Thus, ASD fails again by Theorem 6.3.3. □

Remark 6.3.5

It is interesting to compare the argument in Example 6.3.4 with the algebraic geometry arguments later in Chapter 7 (see Proposition 7.6.5). Both arguments involve constructing commutative subalgebras of $M_n(\mathbb{C})$ of dimension greater than n . But the conclusions are subtly different. Algebraic geometry implies that *some* four commuting $n \times n$ matrices must therefore fail ASD, whereas our argument here says these *specific* four matrices fail ASD. □

It is known (via the so-called Wedderburn–Artin theorem for rings) that $A_1, A_2, \dots, A_k \in M_n(\mathbb{C})$ are simultaneously diagonalizable if and only if the subalgebra $\mathbb{C}[A_1, A_2, \dots, A_k]$ is commutative and contains no nonzero nilpotent matrices. Thus, ASD is equivalent to getting perturbations $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k$ that generate a commutative subalgebra without nonzero nilpotents. But how this helps is not clear.

6.4 REDUCTION TO THE NILPOTENT CASE

Many problems in linear algebra reduce to the nilpotent case, and the ASD one is no exception. But what is a little different with the ASD method, as

we will see in the next section, is that we may start off with commuting nilpotent matrices by our reduction principle, but subsequent perturbations will usually produce non-nilpotent commuting matrices. Then we have to invoke the nilpotent reduction principle again. This may continue in a series of steps, going from nilpotent to non-nilpotent and back to nilpotent. In fact, results in Chapter 7 (see Theorem 7.10.5) strongly suggest that, even starting with commuting nilpotent ASD matrices, sometimes it may be impossible to do perturbations entirely within the class of nilpotent matrices in order to reach commuting nilpotent matrices where one of them is 1-regular.¹² On the other hand, ASD is equivalent to this property within the class of commuting matrices by Proposition 6.2.5.

Here is our reduction principle:

Proposition 6.4.1 (ASD Reduction Principle)

Suppose A_1, \dots, A_k are commuting $n \times n$ complex matrices. Then there exists an invertible matrix C such that $C^{-1}A_1C, \dots, C^{-1}A_kC$ are block diagonal matrices with matching block structures and each diagonal block has only a single eigenvalue (ignoring multiplicities). That is, there is a partition $n = n_1 + \dots + n_r$ of n such that

$$C^{-1}A_iC = \text{diag}(B_{i1}, B_{i2}, \dots, B_{ir}) = \begin{bmatrix} B_{i1} & & & & \\ & B_{i2} & & & \\ & & B_{i3} & & \\ & & & \ddots & \\ & & & & B_{ir} \end{bmatrix}$$

where each B_{ij} is an $n_j \times n_j$ matrix having only a single eigenvalue λ_{ij} for $i = 1, \dots, k$ and $j = 1, \dots, r$. Let $N_{ij} = B_{ij} - \lambda_{ij}I$, a nilpotent matrix. If for $j = 1, 2, \dots, r$, the commuting nilpotent $n_j \times n_j$ matrices $N_{1j}, N_{2j}, \dots, N_{kj}$ are ASD, then A_1, A_2, \dots, A_k are also ASD.

Proof

The production of C and the B_{ij} has already been covered in Proposition 5.1.1 of Chapter 5. Now suppose N_{1j}, \dots, N_{kj} are ASD for $j = 1, \dots, r$. Then clearly $B_{1j}, B_{2j}, \dots, B_{kj}$ are also ASD. Let $\epsilon > 0$. For $j = 1, \dots, r$ choose simultaneously

12. In the terminology and notation of Chapter 7, this will happen if for some n , the variety $\mathcal{C}(3, n)$ of commuting triples of $n \times n$ matrices is irreducible but the variety $\mathcal{CN}(3, n)$ of commuting triples of $n \times n$ nilpotent matrices is reducible.

diagonalizable $n_j \times n_j$ matrices $\bar{B}_{1j}, \bar{B}_{2j}, \dots, \bar{B}_{kj}$ such that

$$\|\bar{B}_{ij} - B_{ij}\| < \frac{\epsilon}{n\|C\| \cdot \|C^{-1}\|}$$

for all i, j . Let

$$\bar{B}_i = \begin{bmatrix} \bar{B}_{i1} & & & & \\ & \bar{B}_{i2} & & & \\ & & \bar{B}_{i3} & & \\ & & & \ddots & \\ & & & & \ddots \\ & & & & & \bar{B}_{ir} \end{bmatrix}$$

and let $\bar{A}_i = C\bar{B}_iC^{-1}$ for $i = 1, \dots, k$. Then $\bar{B}_1, \dots, \bar{B}_k$ are simultaneously diagonalizable, whence so are $\bar{A}_1, \dots, \bar{A}_k$. Now, using properties of the norm,

$$\begin{aligned} \|\bar{A}_i - A_i\| &= \|C\bar{B}_iC^{-1} - C(C^{-1}A_iC)C^{-1}\| \\ &= \|C(\bar{B}_i - C^{-1}A_iC)C^{-1}\| \\ &\leq \|C\| \cdot \|C^{-1}\| \cdot \|\bar{B}_i - C^{-1}A_iC\| \\ &\leq \|C\| \cdot \|C^{-1}\| \sum_{j=1}^r \|\bar{B}_{ij} - B_{ij}\| \\ &< \|C\| \cdot \|C^{-1}\| \cdot \frac{r\epsilon}{n\|C\| \cdot \|C^{-1}\|} \\ &\leq \epsilon. \end{aligned}$$

This demonstrates that A_1, \dots, A_k are ASD. □

Notice that the splitting in Proposition 6.4.1 will be nontrivial ($r > 1$) if one of A_1, \dots, A_k has at least two distinct eigenvalues. When this does occur, it kicks in a natural induction on the smaller-sized block diagonal matrices for establishing the ASD and related properties.

6.5 SPLITTINGS INDUCED BY EPSILON PERTURBATIONS

Our reduction principle in Proposition 6.4.1 suggests a strategy for establishing ASD for various classes of k -tuples of commuting $n \times n$ matrices A_1, A_2, \dots, A_k . To begin with we can assume the matrices are nilpotent. After a similarity transformation, we can by Theorem 2.3.5 also assume that A_1 is a nilpotent Weyl matrix and A_2, A_3, \dots, A_k are strictly upper triangular matrices.

The Strategy. Given an $\epsilon > 0$, find commuting ϵ -perturbations $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k$ of A_1, A_2, \dots, A_k such that one of the \bar{A}_i has two distinct eigenvalues. The reduction principle applied to $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k$ then gives a nontrivial block diagonal splitting of the \bar{A}_i and, provided the corresponding commuting blocks are all within our chosen class of matrices, we can use induction on the size of the matrices to assume ASD for the various k commuting blocks, and deduce ASD for $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k$. Of course, this yields ASD for A_1, A_2, \dots, A_k as well.

Let us say that an ϵ -perturbation \bar{A} of an $n \times n$ matrix A is **k-correctable** if given any collection $A_1 = A, A_2, \dots, A_k$ of k commuting matrices including A , there are ϵ -perturbations $\bar{A}_2, \dots, \bar{A}_k$ of A_2, \dots, A_k such that $\bar{A}_1 = \bar{A}, \bar{A}_2, \dots, \bar{A}_k$ still commute. Here, by an ϵ -perturbation \bar{X} of a matrix X we are only requiring that $\|\bar{X} - X\| < c\epsilon$ for some constant c , which may depend on X but is independent of ϵ . Formulating ϵ -perturbations without this broader interpretation would lead to cumbersome details later. But our description is still a little loose here.¹³ Invariably, however, we have in mind not just one ϵ -perturbation \bar{A} of the given matrix A for a fixed $\epsilon > 0$ but a whole family $\mathcal{P} = \{\bar{A}(\epsilon) : \epsilon \in \mathbb{R}^+\}$ of perturbations of A , one for each positive ϵ . Now the notion of \mathcal{P} being a k -correctable family of perturbations can be made quite precise. Given a collection $A_1 = A, A_2, \dots, A_k$ of k commuting matrices including A , we require the existence of a real number c such that for all sufficiently small $\epsilon > 0$, there are commuting matrices $\bar{A}_1 = \bar{A}(\epsilon), \bar{A}_2, \dots, \bar{A}_k$ such that

$$\|\bar{A}_i - A_i\| < c\epsilon \text{ for } i = 1, 2, \dots, k.$$

By the standard trick in analysis, we can assume $c = 1$ by replacing \mathcal{P} by the family $\{\bar{A}(\epsilon/c) : \epsilon \in \mathbb{R}^+\}$. Having restored rigor, we now lapse back into our earlier less formal mode.

If A is 1-regular, then every ϵ -perturbation \bar{A} of A is k -correctable for all k . This follows from the argument in Example 6.2.3 because the only matrices that commute with a 1-regular matrix are polynomials in that matrix (Proposition 3.2.4). Outside this 1-regular case, nontrivial correctable perturbations are not easy to spot. However, they play a useful role later in certain cases, in the implementation of our above strategy for establishing the ASD property of commuting nilpotent matrices A_1, \dots, A_k . There we make a k -correctable ϵ -perturbation of A_1 that produces two eigenvalues 0 and ϵ . We then split the matching commuting perturbed matrices $\bar{A}_1, \dots, \bar{A}_k$ using Proposition 6.4.1 and repeat the argument (inductively) on smaller nilpotent matrices. The following proposition warns us, however,

13. And, to quote a World War II caution, "Loose lips sink ships."

that we cannot expect to always have the ϵ -eigenspaces 1-dimensional at each step.

Proposition 6.5.1

Suppose A is an $n \times n$ matrix that is not 1-regular. Then A cannot be perturbed to a diagonalizable matrix by a series of n arbitrarily small 2-correctable perturbations that introduce one new eigenvalue of algebraic multiplicity one at each stage.

Proof

Since A is not 1-regular, $\dim \mathcal{C}(A) > n$ by Propositions 3.2.4, 3.1.1, and the Frobenius Formula 3.1.3. (Recall that we denote the centralizer of a square matrix A by $\mathcal{C}(A)$.) Let $\epsilon > 0$. Suppose B_1, \dots, B_n are successive ϵ -perturbations of A such that B_1 is a 2-correctable perturbation of A , B_i is a 2-correctable perturbation of B_{i-1} for $i = 2, \dots, n$, and B_n has n distinct eigenvalues. Let $\{C_1, \dots, C_m\}$ be a basis for $\mathcal{C}(A)$. By Lemma 6.3.2 we can arrange the choice of ϵ so that any ϵ -perturbations of C_1, \dots, C_m will preserve their linear independence. Since B_1 is a 2-correctable perturbation of A , there are ϵ -perturbations $\bar{C}_1, \dots, \bar{C}_m$ of C_1, \dots, C_m such that $\bar{C}_1, \dots, \bar{C}_m$ centralize B_1 . Hence, $\dim \mathcal{C}(A) \leq \dim \mathcal{C}(B_1)$. Repeating this argument n times, we obtain

$$\dim \mathcal{C}(A) \leq \dim \mathcal{C}(B_1) \leq \dim \mathcal{C}(B_2) \leq \dots \leq \dim \mathcal{C}(B_n),$$

whence $\dim \mathcal{C}(B_n) > n$. But B_n has n distinct eigenvalues and so $\dim \mathcal{C}(B_n) = n$ by Proposition 3.2.4. This contradiction establishes the proposition. \square

Remark 6.5.2

By the argument in Example 6.2.2, every square matrix can be perturbed by a series of n arbitrarily small changes to a diagonalizable matrix by introducing a new eigenvalue of multiplicity 1 at each step. (The same is true of any collection of ASD matrices.) Proposition 6.5.1 says that, in general, not all these perturbations are 2-correctable. Interesting. \square

The question of whether a perturbation is 2-correctable is a very natural one, for the following reason. In the context of our strategy, when we attempt to perturb the Weyr matrix $W = A_1$ in our list A_1, A_2, \dots, A_k of commuting matrices, do we have to look in advance at the form of the other A_i ? Well, unless the proposed perturbation of W to \bar{W} is 2-correctable, we *must* bear in mind the other A_i . Otherwise we are doomed to failure in general—we won't even be able to get a perturbation \bar{A}_2 of A_2 that commutes with \bar{W} , let alone perturbations \bar{A}_i of all the A_i that commute with \bar{W} and with each other. Of course, even if \bar{W} is 2-correctable, we are still not out of the woods if $k > 2$. So the question

of 2-correctability is one of the first questions one asks about any proposed perturbation.

Our next proposition records a strong necessary condition for 2-correctability in terms of dimensions of centralizers. For small matrices, this condition can often be mentally checked using our formula in Proposition 3.2.2 for computing the dimension of the centralizer of a nilpotent Weyr matrix, along with the splitting of the centralizer of a matrix with more than one eigenvalue described in Proposition 3.1.1.

Proposition 6.5.3

Let $A \in M_n(\mathbb{C})$.

- (1) For all sufficiently small ϵ , if \bar{A} is an ϵ -perturbation of A , then $\dim \mathcal{C}(\bar{A}) \leq \dim \mathcal{C}(A)$.
- (2) For all sufficiently small ϵ , if \bar{A} is a 2-correctable ϵ -perturbation of A , then $\dim \mathcal{C}(\bar{A}) = \dim \mathcal{C}(A)$.

Proof

(1) Let $p = \dim \mathcal{C}(A)$ and $q = n^2 - p$. Choose a complementary subspace of $\mathcal{C}(A)$ in $M_n(\mathbb{C})$ generated by independent B_1, B_2, \dots, B_q , that is,

$$M_n(\mathbb{C}) = \mathcal{C}(A) \oplus \langle B_1 \rangle \oplus \langle B_2 \rangle \oplus \dots \oplus \langle B_q \rangle.$$

Thinking of the commutator mapping $X \mapsto [A, X] = AX - XA$ as a linear transformation of $M_n(\mathbb{C})$ whose kernel is $\mathcal{C}(A)$, we see that the complementary subspace is faithfully mapped, whence $[A, B_1], [A, B_2], \dots, [A, B_q]$ are linearly independent. By Lemma 6.3.2, sufficiently small ϵ' -perturbations of these latter matrices will remain independent. Next, for a fixed matrix $B \in M_n(\mathbb{C})$, and thinking this time of the mapping $X \mapsto [X, B] = XB - BX$ as a continuous mapping of $M_n(\mathbb{C})$, we see that we can make $[\bar{X}, B]$ an ϵ' -perturbation of $[X, B]$ if \bar{X} is an ϵ -perturbation of X for all sufficiently small ϵ . Therefore, for all sufficiently small ϵ , if \bar{A} is an ϵ -perturbation of A , then

$$[\bar{A}, B_1], [\bar{A}, B_2], \dots, [\bar{A}, B_q] \text{ are linearly independent.}$$

This implies

$$\mathcal{C}(\bar{A}) \cap (\langle B_1 \rangle \oplus \langle B_2 \rangle \oplus \dots \oplus \langle B_q \rangle) = 0.$$

For if $C = b_1 B_1 + b_2 B_2 + \dots + b_q B_q \in \mathcal{C}(\bar{A})$, then $0 = [\bar{A}, C] = b_1 [\bar{A}, B_1] + b_2 [\bar{A}, B_2] + \dots + b_q [\bar{A}, B_q]$, which implies $b_1 = b_2 = \dots = b_q = 0$ and so $C = 0$. Thus, the codimension of $\mathcal{C}(\bar{A})$ in $M_n(\mathbb{C})$ is at least q , whence

$$\dim \mathcal{C}(\bar{A}) \leq n^2 - q = p = \dim \mathcal{C}(A).$$

This establishes (1).

(2) By exactly the same proof as in Proposition 6.5.1, if \bar{A} is a 2-correctable ϵ -perturbation of A for a sufficiently small ϵ , then $\dim \mathcal{C}(\bar{A}) \geq \dim(\mathcal{C}(A))$. Therefore from (1), we have $\dim \mathcal{C}(\bar{A}) = \dim \mathcal{C}(A)$. \square

Remark 6.5.4

The converse of (2) seems plausible, as does its failure! We have been unable to resolve the issue, despite some serious attempts at counter-examples. \square

Example 6.5.5

Consider the nilpotent Weyr matrix

$$A = \left[\begin{array}{ccc|cc|c} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline & & & 0 & 0 & 1 \\ & & & 0 & 0 & 0 \\ & & & & & 0 \end{array} \right]$$

and the following two ϵ -perturbations of A :

$$B = \left[\begin{array}{ccc|cc|c} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline & & & 0 & 0 & 1 \\ & & & 0 & 0 & 0 \\ & & & & & \epsilon \end{array} \right], \quad C = \left[\begin{array}{ccc|cc|c} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \epsilon & 0 & 0 & 0 \\ \hline & & & 0 & 0 & 1 \\ & & & 0 & 0 & 0 \\ & & & & & 0 \end{array} \right].$$

Using our formula in Proposition 3.2.2 for computing the dimension of centralizers, after noting that A has Weyr structure $(3, 2, 1)$, we obtain $\dim \mathcal{C}(A) = 3^2 + 2^2 + 1^2 = 14$. Now B has two eigenvalues 0 and ϵ with algebraic multiplicities 5 and 1, respectively. By the Corollary 1.5.4 to the generalized eigenspace decomposition, B is similar to $\text{diag}(B_1, B_2)$ where B_1 is a 5×5 nilpotent matrix and B_2 is the 1×1 matrix $[\epsilon]$. Since B has rank 3, necessarily B_1 has rank 2 and hence nullity 3. Also the nilpotency index of B_1 is 2 because the rank of B^2 is 1. Thus, by Proposition 2.2.3 the Weyr structure of B_1 is $(3, 2)$. Hence, $\dim \mathcal{C}(B_1) = 3^2 + 2^2 = 13$. Obviously, $\dim \mathcal{C}(B_2) = 1$. Therefore, by Proposition 3.1.1,

$$\dim \mathcal{C}(B) = \dim \mathcal{C}(B_1) + \dim \mathcal{C}(B_2) = 13 + 1 = 14.$$

Thus, B passes the dimension test $\dim \mathcal{C}(B) = \dim \mathcal{C}(A)$ in Proposition 6.5.3 (2), and so is potentially 2-correctable. We leave it as an exercise to show that B is indeed 2-correctable.

On the other hand, C fails the dimension test. For C is similar to $\text{diag}(C_1, C_2)$ where C_1 is a 5×5 nilpotent matrix and C_2 is the 1×1 matrix $[\epsilon]$. But the nullity of the matrix C_1 is 2 and its nilpotency index is 3 (because C^2 has rank 2 and C^3 has rank 1). Hence, the Weyr structure of C_1 is $(2, 2, 1)$. Now

$$\begin{aligned}\dim \mathcal{C}(C) &= \dim \mathcal{C}(C_1) + \dim \mathcal{C}(C_2) \\ &= (2^2 + 2^2 + 1^2) + 1 = 10 \\ &< \dim \mathcal{C}(A).\end{aligned}$$

Thus, C is not a 2-correctable perturbation of A . One could also check this out directly here by computing the exact form of the matrices that centralize C . That involves much more work. It is easier to compute the dimension of the centralizer than it is to work out exactly what the centralizer is. \square

6.6 THE CENTRALIZER OF ASD MATRICES

We already know two purely algebraic necessary conditions for a collection A_1, A_2, \dots, A_k of complex $n \times n$ matrices to have the ASD property: (1) They must commute, and (2) they generate a subalgebra of dimension at most n . In this section, we establish a third such condition in the form of a nice companion piece for (2). Namely, the centralizer $\mathcal{C}(A_1, A_2, \dots, A_k)$ of ASD matrices must have dimension at least n . Here, by $\mathcal{C}(A_1, A_2, \dots, A_k)$ we mean the subalgebra of $M_n(\mathbb{C})$ consisting of the matrices that commute with all the A_1, A_2, \dots, A_k . We begin with a lemma that generalizes Proposition 6.5.3 (1).

Lemma 6.6.1

Let $A_1, A_2, \dots, A_k \in M_n(\mathbb{C})$. For all sufficiently small $\epsilon > 0$, if $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k$ are ϵ -perturbations of A_1, A_2, \dots, A_k respectively, then

$$\dim \mathcal{C}(\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k) \leq \dim \mathcal{C}(A_1, A_2, \dots, A_k).$$

Proof

We extend the argument used in the proof of Proposition 6.5.3 (1). Let $p = \dim \mathcal{C}(A_1, A_2, \dots, A_k)$ and $q = n^2 - p$. Choose independent $B_1, B_2, \dots, B_q \in M_n(\mathbb{C})$ such that

$$M_n(\mathbb{C}) = \mathcal{C}(A_1, A_2, \dots, A_k) \oplus \langle B_1 \rangle \oplus \langle B_2 \rangle \oplus \dots \oplus \langle B_q \rangle.$$

The mapping $\theta : X \mapsto ([A_1, X], [A_2, X], \dots, [A_k, X])$ is a linear transformation of $M_n(\mathbb{C})$ into the product $M_n(\mathbb{C}) \times M_n(\mathbb{C}) \times \dots \times M_n(\mathbb{C})$ of k copies

of $M_n(\mathbb{C})$, and the kernel of θ is $\mathcal{C}(A_1, A_2, \dots, A_k)$. Therefore the images of B_1, B_2, \dots, B_q under θ are linearly independent, that is, the vectors

$$\begin{aligned} V_1 &= ([A_1, B_1], [A_2, B_1], [A_3, B_1], \dots, [A_k, B_1]) \\ V_2 &= ([A_1, B_2], [A_2, B_2], [A_3, B_2], \dots, [A_k, B_2]) \\ &\vdots \\ V_q &= ([A_1, B_q], [A_2, B_q], [A_3, B_q], \dots, [A_k, B_q]) \end{aligned}$$

are linearly independent. By Lemma 6.3.2, for all sufficiently small perturbations $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k$ of A_1, A_2, \dots, A_k , the vectors

$$\begin{aligned} \bar{V}_1 &= ([\bar{A}_1, B_1], [\bar{A}_2, B_1], [\bar{A}_3, B_1], \dots, [\bar{A}_k, B_1]) \\ \bar{V}_2 &= ([\bar{A}_1, B_2], [\bar{A}_2, B_2], [\bar{A}_3, B_2], \dots, [\bar{A}_k, B_2]) \\ &\vdots \\ \bar{V}_q &= ([\bar{A}_1, B_q], [\bar{A}_2, B_q], [\bar{A}_3, B_q], \dots, [\bar{A}_k, B_q]) \end{aligned}$$

are also independent.

Claim: $\mathcal{C}(\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k) \cap \langle B_1, B_2, \dots, B_q \rangle = 0$.

For suppose $C = b_1 B_1 + b_2 B_2 + \dots + b_q B_q \in \mathcal{C}(\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k)$. Then

$$0 = [\bar{A}_i, C] = b_1 [\bar{A}_i, B_1] + b_2 [\bar{A}_i, B_2] + \dots + b_q [\bar{A}_i, B_q]$$

for $i = 1, 2, \dots, k$. Therefore, $b_1 \bar{V}_1 + b_2 \bar{V}_2 + \dots + b_q \bar{V}_q = 0$, whence from the linear independence of the \bar{V}_i we have $b_1 = b_2 = \dots = b_q = 0$. Hence, $C = 0$, establishing our claim.

Finally, as a consequence of the claim, we have the codimension in $M_n(\mathbb{C})$ of $\mathcal{C}(\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k)$ is at least q and so

$$\dim \mathcal{C}(\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k) \leq n^2 - q = p = \dim \mathcal{C}(A_1, A_2, \dots, A_k).$$

□

Theorem 6.6.2

If complex $n \times n$ matrices A_1, A_2, \dots, A_k can be approximated by simultaneously diagonalizable matrices, then $\dim \mathcal{C}(A_1, A_2, \dots, A_k) \geq n$.

Proof

Assume the ASD property for A_1, A_2, \dots, A_k . Choose $\epsilon > 0$ such that the dimension of the centralizers conclusion of Lemma 6.6.1 holds. Let $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k$ be simultaneously diagonalizable matrices that are ϵ -perturbations of our matrices A_1, A_2, \dots, A_k . Let C be an invertible matrix that conjugates $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k$ to diagonal matrices B_1, B_2, \dots, B_k . All the diagonal matrices centralize B_1, B_2, \dots, B_k and, therefore, $\dim \mathcal{C}(B_1, B_2, \dots, B_k) \geq n$. The dimension of $\mathcal{C}(B_1, B_2, \dots, B_k)$ is unchanged if we replace the B_i by their conjugates under C^{-1} . Therefore, $\dim \mathcal{C}(\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k) \geq n$. By Lemma 6.6.1,

$$\dim \mathcal{C}(A_1, A_2, \dots, A_k) \geq \dim \mathcal{C}(\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k) \geq n,$$

yielding our desired conclusion. \square

In Example 6.3.4, we demonstrated the failure of ASD for certain commuting $n \times n$ matrices A_1, A_2, \dots, A_k by showing $\dim \mathbb{C}[A_1, A_2, \dots, A_k] > n$. Clearly, our new test in Theorem 6.6.2 can't be used here because $\dim \mathcal{C}(A_1, A_2, \dots, A_k) \geq \dim \mathbb{C}[A_1, A_2, \dots, A_k] > n$. But there are other situations where our new test denies ASD but for which the old test can't be used. (Of course, as tools for denying ASD, the two tests are mutually exclusive.) For instance, this is the case if A_1, A_2, \dots, A_k generate a maximal commutative subalgebra of $M_n(\mathbb{C})$ of dimension strictly less than n . For then $\dim \mathcal{C}(A_1, A_2, \dots, A_k) = \dim \mathbb{C}[A_1, A_2, \dots, A_k] < n$. Such examples exist. The first of these was given in 1965, when R. C. Courter constructed a maximal commutative subalgebra of $M_{14}(F)$ (over any field F) of dimension 13.¹⁴ When $F = \mathbb{C}$, any generators of that subalgebra must fail ASD by our Theorem 6.6.2. It would be nice to construct three commuting $n \times n$ complex matrices A_1, A_2, A_3 for small n (but necessarily $n > 8$ as we will see in Section 6.12) such that either $\dim \mathbb{C}[A_1, A_2, A_3] > n$ or $\dim \mathcal{C}(A_1, A_2, A_3) < n$. Such matrices would fail ASD. Although triples of commuting matrices that fail ASD are known to exist (this we establish in Chapter 7), apparently no one has ever come face to face with these beasts.

The example by de Boer and Shekhtman, mentioned in Section 6.3, of 16 commuting 17×17 matrices that generate a 17-dimensional subalgebra but fail ASD, shows that our three known necessary conditions for ASD (in terms of commutativity, dimension of the subalgebra, and dimension of the centralizer) are not sufficient even when combined. We end this section with a corollary to

14. Courter's matrix order 14 is minimal. In the 1990s, numerous authors extended his construction to produce families of order 14 and higher. See the papers by Brown, Brown and Call, and Song.

the two dimension conditions (Theorems 6.3.3 and 6.6.2). Its two-line proof is left for the reader's enjoyment.

Corollary 6.6.3

Any maximal commutative subalgebra of $M_n(\mathbb{C})$ that can be generated by ASD matrices must have dimension exactly n .

6.7 A NICE 2-CORRECTABLE PERTURBATION

Chapter 7 will tell us that ASD fails in general for triples of $n \times n$ commuting matrices. So with the benefit of foresight, we realize that there can be no 3-correctable perturbation of a general $n \times n$ nonzero nilpotent matrix that introduces two distinct eigenvalues. That doesn't preclude such 3-correctable perturbations for special classes of nilpotent matrices. On the positive side of the ledger, every nonzero¹⁵ nilpotent matrix has a 2-correctable ϵ -perturbation that has 0 and ϵ as eigenvalues, as we next demonstrate.¹⁶

Proposition 6.7.1

Suppose J and K are commuting matrices with J nonzero and nilpotent. Let Q be a quasi-inverse¹⁷ for J , that is, $J = JQJ$. (If J is in Jordan or Weyr form, one natural choice for Q is the transpose of J .) Let $E = I - JQ$ and suppose $EQ^m = Q^m$ for some $m > 0$ (e.g., Q nilpotent). Let $\epsilon > 0$ and let $L = \epsilon Q + \epsilon^2 Q^2 + \cdots + \epsilon^m Q^m$. Then:

- (1) *the matrices $\bar{J} = J + \epsilon E$ and $\bar{K} = K + LKE$ commute;*
- (2) *\bar{J} has 0 and ϵ as eigenvalues.*

Thus, \bar{J} is a 2-correctable perturbation of J with two distinct eigenvalues.

Note: $\|L\|$ can be made arbitrarily small and, therefore, since K and E are fixed here, so too can $\|LKE\|$. In other words, \bar{K} is an ϵ -perturbation of K .

Proof

Note the relations

$$(i) EJ = 0; \quad (ii) E^2 = E; \quad (iii) EK = EKE.$$

15. It is important to have *nonzero* because the zero matrix, or any scalar matrix, can't have a 2-correctable perturbation with two distinct eigenvalues. Why?

16. This first appeared in the 2006 paper of O'Meara and Vinsonhaler.

17. "Quasi-inverse" here agrees with the von Neumann regular concept we met earlier in Chapter 4. Some readers may wish to take the Moore-Penrose inverse of J as their quasi-inverse Q .

The third equation follows from $EK - EKE = EK(I - E) = EK(JQ) = E(KJ)Q = E(JK)Q = 0$ using (i). Now,

$$\begin{aligned}\bar{J}\bar{K} &= JK + JLKE + \epsilon EK + \epsilon ELKE \\ &= JK + J\left(\sum_{i=1}^m \epsilon^i Q^i\right)KE + \epsilon EK + \epsilon E\left(\sum_{i=1}^m \epsilon^i Q^i\right)KE \\ &= JK + \epsilon(EK + JQKE) + \sum_{i=2}^m \epsilon^i [JQ^i KE + EQ^{i-1} KE] \\ &\quad + \epsilon^{m+1} EQ^m KE. \quad (*)\end{aligned}$$

Similarly,

$$\begin{aligned}\bar{K}\bar{J} &= KJ + LKEJ + \epsilon KE + \epsilon LKE^2 \\ &= KJ + \epsilon KE + \epsilon\left(\sum_{i=1}^m \epsilon^i Q^i\right)KE \quad \text{using (i) and (ii)} \\ &= KJ + \epsilon KE + \sum_{i=2}^m \epsilon^i Q^{i-1} KE + \epsilon^{m+1} Q^m KE. \quad (**)\end{aligned}$$

We now compare the expressions (*) and (**). We have $KJ = JK$ by assumption. Moreover, the coefficients of ϵ agree because $EK + JQKE = EK + (I - E)KE = EK + KE - EKE = KE$ by (iii). The ϵ^i terms agree for $i = 2, 3, \dots, m$ because $JQ^i KE + EQ^{i-1} KE = (I - E)Q^{i-1} KE + EQ^{i-1} KE = Q^{i-1} KE$. Finally, the ϵ^{m+1} terms agree because by assumption $EQ^m = Q^m$. Hence, part (1) of the proposition holds.

We can see that ϵ is an eigenvalue of \bar{J} since $E[\epsilon I - (J + \epsilon E)] = \epsilon E - EJ - \epsilon E^2 = \epsilon E - 0 - \epsilon E = 0$ shows that $\epsilon I - \bar{J}$ is singular. Also, 0 is an eigenvalue of \bar{J} because if $p > 1$ is the nilpotency index of J , then $(J + \epsilon E)J^{p-1} = J^p + \epsilon EJ^{p-1} = 0 + 0 = 0$, which shows \bar{J} is singular. \square

Example 6.7.2

We illustrate Proposition 6.7.1 with a specific 2-correctable perturbation. Consider the 4×4 nilpotent matrix

$$J = \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & 1 \\ & & & 0 \end{array} \right].$$

Here J is in Weyr form but our illustration would work just as well with the Jordan form. Let Q be the transpose of J . Then Q is a nilpotent quasi-inverse of J with $Q^3 = 0$, $JQ = \text{diag}(1, 0, 1, 0)$, and $E = I - JQ = \text{diag}(0, 1, 0, 1)$. The

2-correctable perturbation \bar{J} described in Proposition 6.7.1 is

$$\bar{J} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & \epsilon & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \epsilon \end{bmatrix}.$$

By Proposition 3.2.1, a general matrix that commutes with J takes the form

$$K = \left[\begin{array}{cc|cc} a & b & d & e \\ 0 & c & 0 & f \\ \hline & & a & d \\ & & \hline & & & a \end{array} \right].$$

The corresponding perturbed matrix, described in the proposition, which commutes with \bar{J} is

$$\begin{aligned} \bar{K} &= K + (\epsilon Q + \epsilon^2 Q^2)KE \\ &= \begin{bmatrix} a & b & d & e \\ 0 & c & 0 & f \\ 0 & 0 & a & d \\ 0 & 0 & 0 & a \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \epsilon & 0 & 0 & 0 \\ \epsilon^2 & 0 & \epsilon & 0 \end{bmatrix} \begin{bmatrix} a & b & d & e \\ 0 & c & 0 & f \\ 0 & 0 & a & d \\ 0 & 0 & 0 & a \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} a & b & d & e \\ 0 & c & 0 & f \\ 0 & \epsilon b & a & d + \epsilon e \\ 0 & \epsilon^2 b & 0 & a + \epsilon^2 e + \epsilon d \end{bmatrix}. \end{aligned}$$

The skeptical reader may wish to directly verify that \bar{J} and \bar{K} do actually commute. And the curious reader may ask how we can uncover such perturbations in the first place. It is difficult, when faced with a large nilpotent J , to make a fairly arbitrary perturbation \bar{J} so as to obtain more than one eigenvalue but then recover commutativity with K in a matching perturbation \bar{K} . One can, however, play round with small matrices, observe a pattern, and then attempt to express the perturbations generally in terms of matrix *equations*, not matrix entries.¹⁸ Notice that, in our

18. To check the equality of two expressions, involving products of matrices, by using a series of matrix equations is to invoke the algebra of matrices. To check directly in terms of matrix entry calculations, or expressions involving matrix units, is often to re-check associativity of matrix multiplication as well!

example, the ϵ -eigenvalue of \bar{J} has (algebraic and geometric) multiplicity 2 (and in general has multiplicity the nullity of J).

Our nilpotent J above is 2-regular and with a nonhomogeneous Weyr structure $(2, 1, 1)$. It will follow from work to come in Section 6.10 (Propositions 6.10.5, 6.10.6, and Remark 6.10.8) that the following perturbation of J

$$\begin{aligned}\bar{J} &= J + \epsilon e_{33} \\ &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \epsilon & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}\end{aligned}$$

is k -correctable for all k . Here ϵ is an eigenvalue of multiplicity 1, which appears to be at odds with Proposition 6.5.1 because J is not 1-regular. In actual fact, there is no conflict. What Proposition 6.5.1 says is that we can't continue with three more 2-correctable perturbations, each introducing a new eigenvalue of multiplicity 1. Otherwise we get a contradiction to $\dim \mathcal{C}(J) = 2^2 + 1^2 + 1^2 = 6 > 4$. The reader can check that $\dim \mathcal{C}(\bar{J}) = 6$, so we have $\dim \mathcal{C}(\bar{J}) = \dim \mathcal{C}(J)$, in accordance with Proposition 6.5.3 (2). On the other hand, subsequent (noncorrectable) perturbations resulting in four distinct eigenvalues need to lower the dimension of the centralizer of the final perturbed matrix to 4. \square

6.8 THE MOTZKIN–TAUSSKY THEOREM

One of the earliest results on the ASD property is the following 1955 theorem of Motzkin and Taussky. We shall give a fuller version of the theorem in Chapter 7, in terms of the irreducibility of a certain algebraic variety over any algebraically closed field.

Theorem 6.8.1 (Motzkin–Tausky)

Every pair A_1, A_2 of complex commuting $n \times n$ matrices has the ASD property.

We shall give two proofs. The first is in essence the original proof by Motzkin and Taussky. Their ϵ -perturbations are very special (they are not 2-correctable) and of a different nature to the perturbations we use later. Our second, and much shorter, proof uses the 2-correctable perturbation of the previous section and is more typical of the arguments to come.

First Proof. Suppose A_1 and A_2 are commuting $n \times n$ complex matrices, which, by the reduction principle 6.4.1, we can assume are nilpotent. If A_1 is 1-regular, then A_2 is already a polynomial in A_1 , so we know A_1, A_2 are ASD by Example 6.2.3. Now suppose A_1 is not 1-regular. Let (m_1, m_2, \dots, m_s) be the Jordan structure of A_1 and note $s > 1$ because A_1 is not 1-regular. The diagonal

matrix $\text{diag}(1, 1, \dots, 1, 0, 0, \dots, 0)$ with m_1 ones followed by $n - m_1$ zeros is a proper idempotent matrix, which centralizes the Jordan form of A_1 . Hence, there is a proper idempotent $E \in M_n(\mathbb{C})$ that centralizes A_1 .

The condition that a matrix (over any algebraically closed field) has only a single eigenvalue can be expressed as a (multivariable) polynomial equation in the entries of the matrix.¹⁹ Therefore, if we regard A_2 and E as fixed matrices, then for a scalar c , the condition that $cA_2 + E$ has two distinct eigenvalues is equivalent to $p(c) \neq 0$ for some fixed (single variable) polynomial $p(x) \in \mathbb{C}[x]$. This polynomial is nonzero²⁰ because $p(0) \neq 0$ (since E has the two distinct eigenvalues 0 and 1). In particular, since nonzero polynomials over a field have only finitely many zeros, we have $p(c) \neq 0$ for all but a finite number of complex numbers c . Hence, for all sufficiently small positive ϵ , we have that $A_2 + \epsilon E = \epsilon((1/\epsilon)A_2 + E)$ has two distinct eigenvalues (but not necessarily 0 and ϵ because E may not commute with A_2). Also $A_2 + \epsilon E$ is an ϵ -perturbation of A_2 which commutes with A_1 , because both A_2 and E commute with A_1 . We now have a proper block diagonal splitting of A_1 and $A_2 + \epsilon E$ by Proposition 6.4.1, whence induction applied to the matching commuting blocks completes the proof. \square

Example 6.8.2

A similar technique cannot work for three commuting nilpotent matrices A_1, A_2, A_3 , namely, attempting to perturb A_3 by ϵE for some proper idempotent $E \in \mathcal{C}(A_1, A_2)$. Such an idempotent may not exist, even when A_1 and A_2 are not 1-regular. For instance, if

$$A_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

then

$$\mathcal{C}(A_1, A_2) = \left\{ \begin{bmatrix} p & 0 & q & r \\ 0 & p & s & t \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{bmatrix} : p, q, r, s, t \in \mathbb{C} \right\},$$

and the only idempotents of that subalgebra are 0 and I .

19. This is spelled out fully in Chapter 7, Example 7.1.11 (ii).

20. A subtle point, often overlooked in these types of arguments. The equivalence of distinct eigenvalues to $p(c) \neq 0$ is all very well, but of no use to us here if the underlying polynomial $p(x)$ is identically zero!

The Motzkin–Taussky argument applied to the commuting pair A_1 and A_2 , and using the proper idempotent $E = \text{diag}(1, 0, 1, 0)$, which centralizes A_1 , results in the perturbations $\bar{A}_1 = A_1$ and

$$\bar{A}_2 = A_2 + \epsilon E = \begin{bmatrix} \epsilon & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \epsilon & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The latter perturbation of A_2 is not even 2-correctable because $\dim \mathcal{C}(A_2) = 3^2 + 1^2 = 10$ (note A_2 has nullity 3 and therefore Weyr structure $(3, 1)$), whereas $\dim \mathcal{C}(\bar{A}_2) = 2^2 + 2^2 = 8$ (note \bar{A}_2 is similar to $\text{diag}(\epsilon, \epsilon, 0, 0)$). But the dimension of the centralizer can't decrease under a 2-correctable perturbation (Proposition 6.5.3). We can also see this lack of 2-correctability directly because the $(1, 4)$ entry of a matrix that centralizes \bar{A}_2 must be zero. Therefore, A_3 (which centralizes A_2) can't be perturbed to a matrix that centralizes \bar{A}_2 .

The upshot of this discussion is that, nice though the Motzkin–Taussky perturbations are, they hold little hope of corraling more than two commuting horses at a time.²¹ On the other hand, our arguments in Sections 6.9 and 6.10 will show that any three commuting matrices, one of which is 2-regular (such as with our three commuting matrices A_1, A_2, A_3 above), do indeed have the ASD property. So there do exist commuting perturbations, one (all) of which has distinct eigenvalues. \square

Second Proof. Again, we can assume A_1 and A_2 are commuting nilpotent matrices with A_1 nonzero. Let $\epsilon > 0$. By Proposition 6.7.1 there is a 2-correctable ϵ -perturbation \bar{A}_1 of A_1 that has 0 and ϵ as eigenvalues. Let \bar{A}_2 be a matching perturbation of A_2 that commutes with \bar{A}_1 . Again Proposition 6.4.1 provides a block diagonal splitting of the perturbed matrices \bar{A}_1, \bar{A}_2 and induction finishes off the proof. \square

As a corollary of the Motzkin–Taussky Theorem 6.8.1 and our earlier result that $n \times n$ ASD matrices can't generate more than an n -dimensional subalgebra (Theorem 6.3.3), we obtain a novel proof of Gerstenhaber's Theorem 5.3.2 in the special case of complex matrices.

Corollary 6.8.3 (Gerstenhaber)

Every 2-generated commutative subalgebra of $M_n(\mathbb{C})$ has dimension at most n .

21. Not an "OK Corral" by Frankie Laine's standards.

Proof

Suppose the subalgebra \mathcal{A} of $M_n(\mathbb{C})$ is generated by commuting matrices A_1, A_2 . By Theorem 6.8.1, A_1 and A_2 are ASD and so, by Theorem 6.3.3, $\dim \mathcal{A} \leq n$. \square

Example 6.8.4

To reinforce our strategy outlined in Section 6.5, involving splittings of appropriate perturbations, let us work through the steps in the second proof of the Motzkin–Tausky theorem in the case of the two commuting nilpotent 4×4 matrices

$$A_1 = \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & 1 \\ & & \hline & & & 0 \end{array} \right], \quad A_2 = \left[\begin{array}{cc|cc} 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & -2 \\ \hline & & 0 & -1 \\ & & \hline & & & 0 \end{array} \right]$$

where the first matrix is in Weyr form. Without proper perturbations, these two matrices are certainly not simultaneously diagonalizable because a nonzero nilpotent matrix is never diagonalizable. (A nonzero diagonal matrix is not nilpotent.) Thus, some work is called for. We first perturb A_1, A_2 using the 2-correctable perturbation in Example 6.7.2 (there the matrices are called J, K):

$$\bar{A}_1 = \left[\begin{array}{cccc} 0 & 0 & 1 & 0 \\ 0 & \epsilon & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \epsilon \end{array} \right], \quad \bar{A}_2 = \left[\begin{array}{cccc} 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & -2 \\ 0 & \epsilon & 0 & -1 + 2\epsilon \\ 0 & \epsilon^2 & 0 & -\epsilon + 2\epsilon^2 \end{array} \right].$$

We next look at the block diagonal splittings of these perturbed matrices. The characteristic polynomial of \bar{A}_1 is $x^2(x - \epsilon)^2$. One checks that

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ \epsilon \\ \epsilon^2 \end{bmatrix} \right\}$$

is a basis for \mathbb{C}^4 in which the first two vectors span the null space of \bar{A}_1^2 and the last two span the null space of $(\epsilon I - \bar{A}_1)^2$. The corresponding decomposition into a direct sum of two 2-dimensional subspaces is the generalized eigenspace

decomposition of \bar{A}_1 . Therefore, conjugating by

$$C = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & \epsilon \\ 0 & 0 & 0 & \epsilon^2 \end{bmatrix}, \text{ whose inverse is } C^{-1} = \begin{bmatrix} 1 & 0 & 0 & -1/\epsilon^2 \\ 0 & 0 & 1 & -1/\epsilon \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1/\epsilon^2 \end{bmatrix},$$

yields the block diagonal splittings

$$C^{-1}\bar{A}_1 C = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \epsilon & 0 \\ 0 & 0 & 0 & \epsilon \end{array} \right], \quad C^{-1}\bar{A}_2 C = \left[\begin{array}{cc|cc} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -2\epsilon^2 \\ 0 & 0 & 1 & -\epsilon + 2\epsilon^2 \end{array} \right].$$

Our general argument is to examine the matching diagonal blocks and perturb them to simultaneously diagonalizable matrices, using the Reduction Principle 6.4.1 and further 2-correctable perturbations. That is not necessary in our example because we can see directly how to do this. The lower pair of 2×2 diagonal blocks are already diagonalizable for small ϵ (one is scalar, the other has distinct eigenvalues), so since they also commute they are simultaneously diagonalizable (by Proposition 6.2.6). The upper pair of 2×2 diagonal blocks can be perturbed to simultaneously diagonalizable matrices by the respective perturbations

$$\begin{bmatrix} \epsilon^3 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} -\epsilon^3 & -1 \\ 0 & 0 \end{bmatrix}.$$

These two perturbed matrices have distinct eigenvalues, and they commute (one is the negative of the other), whence are simultaneously diagonalizable. The reason for perturbing by ϵ^3 rather than ϵ here is that we want an ϵ -perturbation of our original matrices when we pull everything back under the inverse conjugation $X \mapsto CXC^{-1}$. Thus, the change $C^{-1}\bar{A}_i C + E_i$ needs to have $\|CE_i C^{-1}\| \leq \epsilon$, which happens when $\|E_i\| \leq \epsilon / (\|C\| \cdot \|C^{-1}\|)$. Since $\|C\|$ is of the order 1, and $\|C^{-1}\|$ is of the order $1/\epsilon^2$, we take $\|E_i\| \leq \epsilon/(1/\epsilon^2) = \epsilon^3$. The inverse conjugation now yields the following perturbations B_1, B_2 of our original pair of matrices A_1, A_2 to simultaneously diagonalizable matrices:

$$B_1 = C \left[\begin{array}{cc|cc} \epsilon^3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \epsilon & 0 \\ 0 & 0 & 0 & \epsilon \end{array} \right] C^{-1} = \begin{bmatrix} \epsilon^3 & 0 & 1 & -\epsilon \\ 0 & \epsilon & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \epsilon \end{bmatrix},$$

$$B_2 = C \left[\begin{array}{cc|cc} -\epsilon^3 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -2\epsilon^2 \\ 0 & 0 & 1 & -\epsilon + 2\epsilon^2 \end{array} \right] C^{-1} = \left[\begin{array}{cccc} -\epsilon^3 & 1 & -1 & 2 + \epsilon \\ 0 & 0 & 0 & -2 \\ 0 & \epsilon & 0 & -1 + 2\epsilon \\ 0 & \epsilon^2 & 0 & -\epsilon + 2\epsilon^2 \end{array} \right].$$

□

6.9 COMMUTING TRIPLES INVOLVING A 2-REGULAR MATRIX

The next step up from the Motzkin and Taussky result for commuting pairs of matrices is the question of ASD for commuting triples A_1, A_2, A_3 of complex $n \times n$ matrices. The results in Chapter 7 will tell us that some further restrictions on the three matrices are necessary for ASD when n is large (even when $n > 28$). What might work? Certainly if one of the matrices is 1-regular, things are fine (Proposition 6.2.5). What about **2-regular**, meaning a matrix whose eigenspaces are at most 2-dimensional? The answer turns out to be “yes,” which we will confirm in this and the following section in Theorem 6.9.1. One can deduce the theorem from a nontrivial algebraic geometry result of Neubauer and Sethuraman in 1999. For from their result, one can deduce that any commuting triple of matrices in which one matrix is 2-regular can be perturbed to a commuting triple in which one matrix is 1-regular. And we know how to proceed from there. Following the treatment in the 2006 paper of O’Meara and Vinsonhaler, we shall give a purely matrix-theoretic proof of the theorem using the Weyr form. The methods may well apply in other situations, such as for 3-regular, but we shall not pursue that here. As always, our primary goal is to illustrate the usefulness of the Weyr form but without (necessarily) giving a definitive account of the particular area for which the Weyr form has some application.

Theorem 6.9.1 (Neubauer–Sethuraman)

The ASD property holds for three commuting matrices over \mathbb{C} if one of them is 2-regular.

If we are presented with commuting matrices A_1, A_2, A_3 with, say, the first matrix 2-regular, we can use a simultaneous similarity transformation to put A_1 in Weyr form W and the other two in upper triangular form (Theorem 2.3.5). It is convenient to label the second and third matrices by K and K' . Clearly if we manage to get ASD for W, K, K' , we have it also for A_1, A_2, A_3 . In terms of the Weyr form W , 2-regular means that in the Weyr structure (n_1, n_2, \dots, n_r) of each of its basic Weyr matrices we have $n_1 \leq 2$. The ASD Reduction Principle 6.4.1 applies because the splitting preserves 2-regularity. Thus, we can assume W is a 2-regular nilpotent Weyr

matrix, and K, K' are strictly upper triangular (because they are nilpotent and upper triangular). Clearly also, there is no loss of generality in assuming W is not 1-regular (because we have the result in the 1-regular case by Proposition 6.2.5). Thus, W has a Weyr structure $(2, 2, \dots, 2, 1, 1, \dots, 1)$. There must be some 2's in this structure but not necessarily any 1's. Without 1's, the Weyr structure is called **homogeneous**. Otherwise, W has a **nonhomogeneous** Weyr structure. For instance $(2, 2, 2, 2)$ is homogeneous whereas $(2, 2, 1, 1, 1)$ is nonhomogeneous. Let us summarize our simplifying assumptions.

Assumptions. W, K, K' are commuting $n \times n$ complex matrices with W a nilpotent Weyr matrix of Weyr structure $(2, 2, \dots, 2, 1, 1, \dots, 1)$, and K, K' strictly upper triangular.

Naturally enough, we will follow the strategy that we outlined in Section 6.5, by perturbing W so as to introduce ϵ as a new eigenvalue (but retaining 2-regularity), and finding matching commuting perturbations of K and K' . Our arguments depend on whether the Weyr structure of W is homogeneous or not. We handle the homogeneous case in this section and the nonhomogeneous case in the one that follows. In both cases, we manage to make ϵ an eigenvalue of the perturbed W of multiplicity one, but the degree of correctability of the perturbation varies according to whether W is homogeneous or not. Our perturbations are applied repeatedly until the final perturbed \overline{W} is diagonalizable. A close analysis of our methods reveals that actually the perturbations used in the nonhomogeneous case are k -correctable for all positive integers k , and the lack of correctability is confined to the homogeneous case, where some perturbations are not even 2-correctable. Even there, however, the perturbation of W has a “limited sort of 3-correctability within upper triangular matrices.” In hindsight, the limited 3-correctability is about the best one could hope for in the homogeneous case, in view of the four commuting 4×4 upper triangular matrices in Example 6.3.4 failing the ASD property. Note there that the subalgebra \mathcal{A} generated by E_1, E_2, E_3, E_4 also has $E_1 + E_4, E_2, E_3, E_4$ as generators, so these new generators must also fail ASD. But the first is a nilpotent Weyr matrix of homogeneous structure $(2, 2)$.

It is important to bear in mind that this section, on its own, won't establish Theorem 6.9.1 in the homogeneous case. The block diagonal splittings of $\overline{W}, \overline{K}, \overline{K}'$, and subsequent nilpotent reductions, that occur after the first perturbation, will generally involve the nonhomogeneous case as well.

We proceed to the homogeneous case, where W is a nilpotent Weyr matrix with Weyr structure $(2, 2, \dots, 2)$ involving, say, r lots of 2's. Note that $n = 2r$ is even and r is the nilpotency index of W . As an $r \times r$ blocked matrix with 2×2

blocks, we have

$$W = \begin{bmatrix} 0 & I & & & & \\ & 0 & I & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 0 & I \\ & & & & & 0 \end{bmatrix}.$$

By Proposition 3.2.1, since K commutes with W , we know K has the form

$$K = \begin{bmatrix} D_0 & D_1 & D_2 & \cdots & D_{r-2} & D_{r-1} \\ & D_0 & D_1 & D_2 & \ddots & D_{r-2} \\ & & D_0 & D_1 & \ddots & \vdots \\ & & & \ddots & \ddots & D_2 \\ & & & & D_0 & D_1 \\ & & & & & D_0 \end{bmatrix}$$

where the D_i are 2×2 matrices with D_0 strictly upper triangular. For a 2×2 matrix D , we will use the notation $[D]$ to denote the $n \times n$ matrix with D 's down the main diagonal and 0's elsewhere. By the shifting effect described in Remark 2.3.1, K can be written uniquely as

$$K = [D_0] + [D_1]W + [D_2]W^2 + \cdots + [D_{r-1}]W^{r-1}.$$

We call this the **W-expansion** of K .²² Note that the coefficients $[D_i]$ commute with W . Similarly, K' can be written as

$$K' = [D'_0] + [D'_1]W + [D'_2]W^2 + \cdots + [D'_{r-1}]W^{r-1}.$$

22. Ring-theoretically expressed, this is just saying that there is a natural algebra isomorphism from the centralizer $\mathcal{C}(W)$ of W to the factor algebra $M_2(F)[t]/(t^r)$ of polynomials in the (commuting) indeterminant t with coefficients from $M_2(F)$, modulo the ideal generated by t^r . Under the isomorphism W becomes t , and K becomes $D_0 + D_1t + D_2t^2 + \cdots + D_{r-1}t^{r-1}$. More generally, when W is a d -regular $n \times n$ nilpotent Weyr matrix with a homogeneous structure, $\mathcal{C}(W) \cong M_d(F)[t]/(t^r)$ where $n = dr$. This abstraction is often the right setting to apply elegant algebraic geometry arguments. However, in other situations, there can be something lost. For instance, in the present chapter, when we construct explicit perturbations of commuting triples of matrices, it is often important not only to know the structure of the centralizer $\mathcal{C}(W)$ of an $n \times n$ nilpotent Weyr matrix W , but also how that centralizer sits inside the full matrix algebra $M_n(F)$. This is because the perturbations will usually take us outside that centralizer.

Thus, K and K' are polynomial expressions in W . These forms are much nicer to work with than the corresponding ones when W is in Jordan form. We can compute the product KK' just as we would the product of two polynomials. Calculations involving the coefficients of our new polynomials are just 2×2 matrix calculations. And we don't have to worry about terms in the product of degree r or more because they are zero. For instance, when $r = 3$,

$$\begin{aligned} & ([D_0] + [D_1]W + [D_2]W^2) ([D'_0] + [D'_1]W + [D'_2]W^2) \\ &= [D_0D'_0] + [D_0D'_1 + D_1D'_0]W + [D_0D'_2 + D_1D'_1 + D_2D'_0]W^2. \end{aligned}$$

Of course, since the product KK' also centralizes W , it must also have the block form above. So we only ever have to compute the first row of blocks in the product in order to know the product exactly. A trivial point, but very useful.

Further simplifications in the forms of K and K' can be achieved using two sorts of operations. We can modify a particular D_i or D'_i , to within similarity, by conjugating W , K , K' by an invertible $[C]$ (this doesn't change W). And we can replace K and K' by any two matrices that, together with W , generate the same subalgebra as W , K , K' . This follows from Proposition 6.3.1.

Lemma 6.9.2 (Standard Form of Generators)

Let F be an algebraically closed field and \mathcal{A} be a commutative subalgebra of $M_n(F)$ containing a nilpotent Weyr matrix W of index r and with homogeneous Weyr structure $(2, 2, \dots, 2)$. If there is a set of k generators for \mathcal{A} that includes W , then there is an integer $h \leq r/2 \leq n/4$ such that \mathcal{A} has a set of k generators $\{W, K, K', \dots\}$ for which K takes the form

$$K = [D_h]W^h + \dots + [D_{r-1}]W^{r-1}$$

and all the other generators from the third onwards take the form

$$K' = [D'_{r-h}]W^{r-h} + \dots + [D'_{r-1}]W^{r-1}.$$

(Note that $KK' = 0$ for all the other generators K' .) We refer to these expressions for K and the K' as being in **standard form**. The standard forms will be nilpotent if the original generators are nilpotent.

Proof

Suppose \mathcal{S} is a set of k generators of \mathcal{A} that includes W . From \mathcal{S} , choose a matrix $K = [D_0] + [D_1]W + \dots + [D_{r-1}]W^{r-1}$ such that its first index h for which D_h is not a scalar matrix is minimal among all such indices over all the generators from \mathcal{S} . We can assume such an index exists, otherwise the algebra \mathcal{A} is generated by W alone, in which case the result is trivial. Modify K by subtracting

$[D_0] + [D_1]W + \cdots + [D_{h-1}]W^{h-1}$, remaining in \mathcal{A} because D_0, D_1, \dots, D_{h-1} are scalar. Similarly, modify the other K' . We still have a set of k generators including W , but now K and K' have the first h coefficients in their W -expansions equal to zero. We can assume $h \leq r/2$, otherwise we could redefine h to be the integer part of $r/2$, set $D_h = 0$, and be finished. Since commutative subalgebras of $M_2(F)$ have dimension at most 2 (see, for example, Theorem 5.4.4), a 2×2 matrix that commutes with the nonscalar matrix D_h must be in the linear span of I and D_h . We can use this and the shifting effect of W under repeated multiplications on K to “clear out” the other generators K' .

To see this, let $K' = [D'_h]W^h + [D'_{h+1}]W^{h+1} + \cdots + [D'_{r-1}]W^{r-1}$ be any of the other generators. If $2h < r$, then by exactly the same argument as in Proposition 3.4.4 (4), D'_h must centralize D_h . Hence,

$$D'_h = aI + bD_h$$

for some scalars a, b . Now replace K' by $K' - aW^h - bK$. This clears out the term $[D'_h]W^h$. We can repeat this argument, successively, on the other coefficients $[D'_s]$ of (the modified) K' for $s < r - h$: D'_s commutes with D_h , so $D'_s = cI + dD_h$ for some scalars c, d and we can subtract $cW^s + dKW^{s-h}$ from K' to clear out $[D'_s]W^s$. Eventually K' will have the form in the lemma. The key point is the shifting effect of W (see Remark 2.3.1), now encapsulated in polynomial multiplication. It is such a wonderful feature of the Weyr form. \square

Remark 6.9.3

In particular, the lemma tells us that if one of the matrices K in \mathcal{A} has its D_0 a nonscalar matrix ($h = 0$), then \mathcal{A} is generated by W and K alone. In fact the proof shows that $\{K^i W^j : i, j = 0, 1, \dots, r - 1\}$ is a vector space basis for \mathcal{A} . In particular, $\dim \mathcal{A} = 2r = n$. The latter dimension is still possible for 3 generators even if all K have a scalar D_0 . For instance, let $W = A_1$, $K = A_2$, $K' = A_3$ where the A_i are as in Example 6.8.2. (These generators are in standard form with $r = 2$, $h = 1$.) \square

We return to establishing ASD for our W, K, K' . Notice that Lemma 6.9.2 now allows us to further assume that K and K' have the standard form, because the ASD property is independent of the generators of the subalgebra $\mathbb{C}[W, K, K']$ (see Proposition 6.3.1). Moreover, in view of Remark 6.9.3 and the Motzkin–Tausky Theorem 6.8.1, we may as well assume $h \geq 1$. (Our proof below still works when $h = 0$, which necessitates $K' = 0$, although case (2) introduces a matrix that is not strictly upper triangular, and for which we need to recheck the case (1) argument.) Henceforth we make these standard form assumptions, and consider two cases:

Case (1): D_h is diagonalizable (and $h \geq 1$). We can conjugate W, K, K' by a block diagonal matrix with a 2×2 matrix P along the diagonal so that the new D_h is diagonal. Next subtract a scalar multiple of W^h so that D_h has the form

$$D_h = \begin{bmatrix} * & 0 \\ 0 & 0 \end{bmatrix}. \text{ We introduce epsilon changes to } W, K, K' \text{ as follows:}$$

Notation 6.9.4

Case (1).

$$Q = W^T \text{ (the transpose of } W \text{)}$$

$$E = e_{22}$$

$$T = e_{22} + e_{44} + \cdots + e_{n-2, n-2}$$

$$K = [D_h]W^h + \cdots + [D_{r-1}]W^{r-1} \text{ with } D_h = \begin{bmatrix} * & 0 \\ 0 & 0 \end{bmatrix}, 1 \leq h \leq r/2$$

$$K' = [D'_{r-h}]W^{r-h} + \cdots + [D'_{r-1}]W^{r-1}$$

$$\overline{W} = W + \epsilon E$$

$$\overline{K} = K - \epsilon QTK$$

$$\overline{K'} = K' - \epsilon QTK'$$

□

A technical lemma establishes useful elementary relationships among the matrices defined above.

Lemma 6.9.5

With notation as in 6.9.4, we have

- (1) $WQT = T$
- (2) $QTW = -E + T + e_{nn}$
- (3) $EQ = 0$
- (4) $KE = 0 = K'E$
- (5) $e_{nn}K = 0 = e_{nn}K'$
- (6) $0 = KK' = K'K = K'QTK = KQTK'$.

Proof

Parts (1) and (2) are straightforward applications of the definitions.

$$(3) EQ = e_{22}Q = 0 \text{ since } Q \text{ has zero second row.}$$

$$(4) Ke_{22} = 0 = K'e_{22}, \text{ since } K \text{ and } K' \text{ have zero second column.}$$

$$(5) e_{nn}K = 0 = e_{nn}K' \text{ because both } K \text{ and } K' \text{ are strictly upper triangular.}$$

(6) From the standard form above, together with the fact that $W^r = 0$, we have $KK' = 0 = K'K$. Note that K begins with $2h$ columns of zeros while K' ends with $n - 2h$ rows of zeros. Now right multiplication by QT shifts the even-numbered

columns of K two to the left and annihilates the other columns. However, the $2h + 2$ column of K is zero because of the form of D_h , whence the first $2h$ columns of KQT are also zero. Hence, $KQTK' = 0$.

Similarly, K' begins with $n - 2h$ columns of zeros and K ends with $2h + 1$ rows of zeros. Next, note that left multiplication by QT shifts the even-numbered rows of K down two and annihilates the others. But the $n - 2h$ row of K is zero because of the form of D_h . Thus, QTK still ends with $2h$ rows of zeros and so $K'QTK = 0$. \square

Proposition 6.9.6

For case (1), and in the notation of 6.9.4, we have that \overline{W} , \overline{K} , \overline{K}' are commuting perturbations of W , K , K' . Moreover, \overline{W} has two eigenvalues, 0 and ϵ , and is 2-regular.

Proof

Using the definitions and identities from Lemma 6.9.5, we have

$$\begin{aligned}\overline{W}\overline{K} &= WK - \epsilon WQTK + \epsilon EK - \epsilon^2 EQTK \\ &= WK - \epsilon TK + \epsilon EK.\end{aligned}$$

On the other hand,

$$\overline{K}\overline{W} = KW + \epsilon KE - \epsilon QTKW - \epsilon^2 QTKE.$$

Again using identities in Lemma 6.9.5, along with $QTKW = QTWK$, we have

$$\overline{K}\overline{W} = KW - \epsilon(-E + T + e_{nn})K.$$

After noting $e_{nn}K = 0$, we see that the expressions for $\overline{W}\overline{K}$ and $\overline{K}\overline{W}$ are equal. To show $\overline{W}K' = K'\overline{W}$, we employ the same argument. Thus, \overline{W} commutes with \overline{K} and \overline{K}' .

To show \overline{K} and \overline{K}' commute, we have

$$\overline{K}\overline{K}' = KK' - \epsilon KQTK' - \epsilon QTKK' + \epsilon^2 QTKQTK' = KK'$$

using the identities in Lemma 6.9.5 (6). Similarly,

$$\overline{K}'\overline{K} = K'K - \epsilon K'QTK - \epsilon QTK'K + \epsilon^2 QTK'QTK = K'K$$

using the identities in Lemma 6.9.5 (6). Since K and K' commute, the proof of this part is complete.

Inasmuch as \overline{W} is upper triangular with zeros down the diagonal except for ϵ in the $(2, 2)$ position, \overline{W} has 0 and ϵ as its eigenvalues. Their geometric multiplicities

are, respectively, 2 and 1, except when $n = 2$, in which case they are both of multiplicity 1. Thus, \overline{W} is 2-regular. \square

Remark. It is easy to check for case (1) that if K'' is another commuting matrix with the same form as K' , then $\overline{K'}$ and $\overline{K''}$ commute whenever the 2×2 block matrices in their expansions satisfy $D'_i = D''_i = 0$ for $i \leq r/2$. Thus, we can introduce epsilon changes to any number of commuting matrices if the beginning indices g in the W -expansions of the matrices other than W and K satisfy $g > r/2$. \square

Case (2): D_h is not diagonalizable (and $h \geq 1$). Here we conjugate with a block diagonal matrix to put D_h into Jordan form. By subtracting a scalar multiple of W^h from K we may assume

$$D_h = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Let L be the block diagonal matrix with repeated 2×2 blocks

$$\begin{bmatrix} 1 & 0 \\ \epsilon & 1 \end{bmatrix}.$$

Then L centralizes W (by Proposition 2.3.3). Moreover, the matrices W , KL , and $K'L$ commute. In fact $KLK'L = 0 = K'CLK$ by the same arguments that show $KK' = 0 = K'K$. Also KL and $K'L$ are nilpotent (strictly upper triangular) because of our assumption that $h \geq 1$. Now KL has for its “ D_h coefficient” the matrix

$$\begin{bmatrix} \epsilon & 1 \\ 0 & 0 \end{bmatrix},$$

which is diagonalizable. Therefore, by case (1) we can obtain commuting perturbations of W , KL , $K'L$ that introduce an ϵ eigenvalue to W . This yields the desired epsilon changes to W , K , K' because KL and $K'L$ are ϵ -perturbations of K and K' , respectively. For instance, $KL = K + KM$ where M is the block diagonal matrix with 2×2 repeated blocks

$$\begin{bmatrix} 0 & 0 \\ \epsilon & 0 \end{bmatrix}.$$

Remark 6.9.7

For the 4×4 nilpotent Weyr matrix W of structure $(2, 2)$, the perturbation of W prescribed in case (1) is

$$\overline{W} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & \epsilon & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

A quick calculation shows that a matrix that centralizes \overline{W} must have a zero $(1, 2)$ entry. Therefore, the matrix

$$A = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

which centralizes W cannot be perturbed to a matrix \overline{A} that centralizes \overline{W} . Thus, our perturbation of W in the 2-regular homogeneous case is not even 2-correctable, unlike the 2-regular nonhomogeneous perturbation in the following section. \square

We have now successfully completed the induction step, according to the strategy in Section 6.5, when W has a homogeneous structure. In the next section we do likewise in the nonhomogeneous case. But first an example illustrating the specifics of our perturbations in the homogeneous case.

Example 6.9.8

Let W be the 6×6 nilpotent Weyr matrix with homogeneous structure $(2, 2, 2)$:

$$W = \left[\begin{array}{cc|cc|cc} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline & & 0 & 0 & 1 & 0 \\ & & 0 & 0 & 0 & 1 \\ \hline & & & & 0 & 0 \\ & & & & 0 & 0 \end{array} \right]$$

Thus, $r = 3$. Let

$$K = \left[\begin{array}{cc|cc|cc} 0 & 0 & 5 & -8 & 0 & 1 \\ 0 & 0 & 4 & -7 & 2 & 3 \\ \hline & & 0 & 0 & 5 & -8 \\ & & 0 & 0 & 4 & -7 \\ \hline & & & & 0 & 0 \\ & & & & 0 & 0 \end{array} \right], \quad K' = \left[\begin{array}{cc|cc|cc} 0 & 0 & -3 & 8 & -1 & 1 \\ 0 & 0 & -4 & 9 & 4 & -2 \\ \hline & & 0 & 0 & -3 & 8 \\ & & 0 & 0 & -4 & 9 \\ \hline & & & & 0 & 0 \\ & & & & 0 & 0 \end{array} \right].$$

Then W, K, K' commute. We wish to perturb them to a commuting triple for which 0 and ϵ are eigenvalues of the perturbed W . We can reach standard form for K and K' by leaving K unchanged and subtracting $2W - K$ from K' (since $D'_1 = 2I - D_1$ here) : replace K' by

$$K'_1 = \left[\begin{array}{cc|cc|cc} 0 & 0 & 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & 0 & 6 & 1 \\ \hline & & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 0 & 0 \\ \hline & & & & 0 & 0 \\ & & & & 0 & 0 \end{array} \right].$$

Note $h = 1$ in the standard form. Since

$$D_1 = \begin{bmatrix} 5 & -8 \\ 4 & -7 \end{bmatrix}$$

is diagonalizable, we are in case (1). We can diagonalize D_1 by

$$\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 5 & -8 \\ 4 & -7 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}.$$

Thus, if we conjugate W, K, K'_1 by the block diagonal matrix

$$Y = \left[\begin{array}{cc|cc|cc} 2 & 1 & & & & \\ 1 & 1 & & & & \\ \hline & & 2 & 1 & & \\ & & 1 & 1 & & \\ \hline & & & & 2 & 1 \\ & & & & 1 & 1 \end{array} \right],$$

we finish up with the commuting triple W, K_1, K'_2 where

$$K_1 = \left[\begin{array}{cc|cc|cc} 0 & 0 & 1 & 0 & -6 & -4 \\ 0 & 0 & 0 & -3 & 13 & 9 \\ \hline & & 0 & 0 & 1 & 0 \\ & & 0 & 0 & 0 & -3 \\ \hline & & & & 0 & 0 \\ & & & & 0 & 0 \end{array} \right], \quad K'_2 = \left[\begin{array}{cc|cc|cc} 0 & 0 & 0 & 0 & -13 & -6 \\ 0 & 0 & 0 & 0 & 26 & 13 \\ \hline & & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 0 & 0 \\ \hline & & & & 0 & 0 \\ & & & & 0 & 0 \end{array} \right].$$

The new D_1 is $\text{diag}(1, -3)$ so we subtract $-3W$ from K_1 to get a further modification

$$K_2 = \left[\begin{array}{cc|cc|cc} 0 & 0 & 4 & 0 & -6 & -4 \\ 0 & 0 & 0 & 0 & 13 & 9 \\ \hline & & 0 & 0 & 4 & 0 \\ & & 0 & 0 & 0 & 0 \\ \hline & & & & 0 & 0 \\ & & & & 0 & 0 \end{array} \right].$$

Now we have modified our original commuting triple W, K, K' to the commuting triple W, K_2, K'_2 for which ASD is equivalent, but to which our case (1) perturbation recipe in 6.9.4 applies. Its ingredients produce the perturbations

$$\overline{W} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \epsilon & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\overline{K}_2 = \begin{bmatrix} 0 & 0 & 4 & 0 & -6 & -4 \\ 0 & 0 & 0 & 0 & 13 & 9 \\ 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & -13\epsilon & -9\epsilon \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \overline{K}'_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & -13 & -6 \\ 0 & 0 & 0 & 0 & 26 & 13 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -26\epsilon & -13\epsilon \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

□

Remark 6.9.9

Working back through the relationships, we have

$$\begin{aligned} K &= YK_2Y^{-1} - 3YWY^{-1}, \\ K' &= YK'_2Y^{-1} + 2W - K, \\ W &= YWY^{-1}. \end{aligned}$$

Therefore, if we wish to, we can also give commuting perturbations $\overline{W}, \overline{K}, \overline{K}'$ of our original matrices by taking

$$\begin{aligned} \overline{K} &= Y\overline{K}_2Y^{-1} - 3Y\overline{W}Y^{-1}, \\ \overline{K}' &= Y\overline{K}'_2Y^{-1} + 2Y\overline{W}Y^{-1} - \overline{K}. \end{aligned}$$

□

Remark 6.9.10

In the above example, the subalgebra $\mathcal{A} = \mathbb{C}[W, K, K']$ is genuinely 3-generated. It can't be generated by two matrices. So we can't deduce ASD for W, K, K' through the Motzkin–Tausky Theorem 6.8.1 and Proposition 6.3.1. We won't go through the details of this, other than to say (i) $\dim \mathcal{A} = 6$ (of course, it can be at most 6 by Theorem 6.3.3 because the generators have ASD) but (ii) a 2-generated subalgebra $\mathcal{B} = \mathbb{C}[A, B]$ of \mathcal{A} has dimension at most 5. A good way of establishing (ii) is to show that, if $\mathcal{B} = \mathcal{A}$, then A can be assumed to be a nilpotent Weyr matrix of homogeneous structure $(2, 2, 2)$, relative to which the first leading edge dimension of \mathcal{B} is 1. By Lemma 5.4.1, the second and third leading edge dimensions of \mathcal{B} (relative to A) are at most 2. Thus, $\dim \mathcal{B} \leq 1 + 2 + 2 = 5$. All this makes for an instructive exercise. \square

6.10 THE 2-REGULAR NONHOMOGENEOUS CASE

In this section, we complete the proof of Theorem 6.9.1 by considering the case of a commuting triple W, K, K' of $n \times n$ matrices for which W is a nilpotent Weyr matrix of nonhomogeneous Weyr structure $(2, 2, \dots, 2, 1, 1, \dots, 1)$, and K, K' are strictly upper triangular. Let s be the number of 2's in the structure of W and let $t = 2s$. As a point of reference when visualizing the matrices, t is the size of the submatrix (upper left corner) involving the 2×2 blocks. (The rest of the blocks are either 2×1 or 1×1 .) Again, by the now well-used centralizing result, we know the forms of K, K' as block upper triangular matrices of the same block structure as W . Rather than display these generally, we use an example to illustrate our arguments. As is often the case in mathematics, the concrete example contains the essence of the general situation.

Example 6.10.1

Let

$$W = \left[\begin{array}{cc|cc|cc} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline & & 0 & 0 & 1 & 0 \\ & & 0 & 0 & 0 & 0 \\ & & & & 0 & 1 \\ & & & & & 0 \end{array} \right]$$

be the nilpotent Weyr matrix of structure $(2, 2, 1, 1)$. Thus, $s = 2$ and $t = 4$. Then K and K' take the form

$$K = \left[\begin{array}{cc|cc|cc} 0 & a & b & c & e & g \\ 0 & 0 & 0 & d & f & h \\ \hline & & 0 & a & b & e \\ & & 0 & 0 & 0 & f \\ & & & & 0 & b \\ & & & & & 0 \end{array} \right], \quad K' = \left[\begin{array}{cc|cc|cc} 0 & a' & b' & c' & e' & g' \\ 0 & 0 & 0 & d' & f' & h' \\ \hline & & 0 & a' & b' & e' \\ & & 0 & 0 & 0 & f' \\ & & & & 0 & b' \\ & & & & & 0 \end{array} \right].$$

Of course, extra conditions are needed on the entries of K and K' for them to commute. \square

Just as in the homogeneous case, we need to “condition” K and K' prior to introducing the commuting perturbations. The freedom to do this is the same: if we haven’t altered the subalgebra generated by the three matrices, then ASD isn’t altered either. The clearing out of K, K' in the nonhomogeneous case, however, is much simpler and only involves subtracting various scalar multiples of powers of W . We will demonstrate this for the above example, after which the pattern should be clear. We replace K and K' by, respectively,

$$K - bW - eW^2 - gW^3 = \left[\begin{array}{cc|cc|cc} 0 & a & 0 & c & 0 & 0 \\ 0 & 0 & 0 & d_1 & f & h \\ \hline & & 0 & a & 0 & 0 \\ & & 0 & 0 & 0 & f \\ & & & & 0 & 0 \\ & & & & & 0 \end{array} \right],$$

$$K' - b'W - e'W^2 - g'W^3 = \left[\begin{array}{cc|cc|cc} 0 & a' & 0 & c' & 0 & 0 \\ 0 & 0 & 0 & d'_1 & f' & h' \\ \hline & & 0 & a' & 0 & 0 \\ & & 0 & 0 & 0 & f' \\ & & & & 0 & 0 \\ & & & & & 0 \end{array} \right],$$

where $d_1 = d - b$ and $d'_1 = d' - b'$. Remember that in these calculations, since we are staying inside the centralizer of W , we only have to keep track of the changes to the first row of blocks; the rest are then completely determined. Also remember what the powers of W look like, and their effect on a matrix under right multiplication. Our clearing out is removing the $(1, 1)$ entry in each of the blocks of K and K' . It is clear that we can do this in general. Henceforth, we assume the following:

Cleared Out Assumption. K and K' have a zero $(1, 1)$ entry in each of their blocks.

Here, as we can quickly check, are three useful consequences of our cleared out assumption. For convenience, we also record property (4), which does

not rely on clearing out, simply on the fact that K and K' centralize W (Proposition 3.2.1).

Lemma 6.10.2

- (1) Rows $t + 1, t + 2, \dots, n$ of K and K' are zero.
- (2) For all odd i and j , the (i, j) entries of K and K' are zero.
- (3) If i is odd and $j > t$, then the (i, j) entries of K and K' are zero.
- (4) The (i, j) entry of K equals its $(i + 2, j + 2)$ entry whenever $i < j \leq t - 1$.
The same applies to K' .

We are now ready to introduce our proposed perturbations $\overline{W}, \overline{K}, \overline{K}'$ of the commuting W, K, K' in the nonhomogeneous case under the above cleared out assumption.

Notation 6.10.3

$$Q = W^T \text{ (the transpose of } W\text{)}$$

$$E = e_{t+1, t+1}$$

$$S = e_{11} + e_{33} + \dots + e_{t+1, t+1}$$

$$\overline{W} = W + \epsilon E$$

$$\overline{K} = K - \epsilon KSQ$$

$$\overline{K}' = K' - \epsilon K'SQ$$

It is clear from the triangular form of \overline{W} that it has 0 and ϵ as eigenvalues. Our remaining goal is to show that the epsilon changes introduced in Notation 6.10.3 do not destroy commutativity. The first lemma establishes some basic relationships among our matrices. The equalities follow from direct computations.

Lemma 6.10.4

In the notation of 6.10.3, if K and K' are cleared out, then:

- (1) $SKS = 0$
- (2) $E = 0$
- (3) $QW = I - e_{11} - e_{22}$
- (4) $Ke_{11} = 0$
- (5) $SQ = SQS$
- (6) $SQE = 0$
- (7) $WSQ = S - E$

Moreover, the same identities hold if K is replaced by K' .

Proof

(1) Multiplying on the left and right by S picks out odd rows and columns of K in the top left $(t + 1) \times (t + 1)$ corner and sets all other entries to zero. By Lemma 6.10.2 (2), we must therefore have $SKS = 0$.

(2) Since EK has nonzero entries only in the $t + 1$ row, this matrix is zero by Lemma 6.10.2 (1).

(3) Right multiplication by W shifts columns 1 to $t - 1$ over to the right by 2, kills column t , then moves columns $t + 1, \dots, n - 1$ over one, and finally kills column n . Viewing the product QW this way gives (3). Alternatively, we can note that left multiplication by Q shifts rows 1 to $t - 1$ down two, kills row t , then shifts rows $t + 1, \dots, n - 1$ down one, and finally kills row n .

(4) Since K is strictly upper triangular, its first column is zero.

(5), (6) The matrix SQ has (i, j) entries that are nonzero only if i, j are odd with $i = j + 2 \leq t + 1$. It is immediate that $SQS = SQ$ and $SQE = 0$.

(7) Multiplication of S on the left by W shifts rows up two (with the top two rows killed), while multiplication of WS by Q on the right shifts columns two to the left.

The last sentence of the lemma is clear from the identical form for K and K' . \square

Proposition 6.10.5

In the notation of 6.10.3, and under the cleared out assumption, $\overline{W} \overline{K} = \overline{K} \overline{W}$ and $\overline{W} \overline{K}' = \overline{K}' \overline{W}$.

Proof

It suffices to prove only the first equality. First, by definition,

$$\overline{W} \overline{K} = WK - \epsilon WKSQ + \epsilon EK - \epsilon^2 EKSQ.$$

We can use (2) of Lemma 6.10.4 to eliminate ϵEK and $\epsilon^2 EKSQ$. Then

$$\begin{aligned} \overline{W} \overline{K} &= WK - \epsilon WKSQ \\ &= WK - \epsilon KWSQ \\ &= WK - \epsilon K(S - E) \\ &= WK - \epsilon KS + \epsilon KE, \end{aligned}$$

by 6.10.4 (7). Next

$$\overline{K} \overline{W} = KW - \epsilon KSQW + \epsilon KE - \epsilon^2 KSQE.$$

Noting that $Se_{11} = e_{11}$ and $Se_{22} = 0$, we can now use (3), (6), and (4) of Lemma 6.10.4 to write

$$\begin{aligned}\overline{K} \overline{W} &= KW - \epsilon KS(I - e_{11} - e_{22}) + \epsilon KE \\ &= KW - \epsilon KS + \epsilon KE,\end{aligned}$$

establishing commutativity, since $WK = KW$. □

So far things have been quite straightforward in the nonhomogeneous case. But to establish commutativity of \overline{K} and \overline{K}' requires a careful argument.

Proposition 6.10.6

In the notation of 6.10.3, if W, K, K' commute (with K and K' cleared out), then $\overline{K} \overline{K}' = \overline{K}' \overline{K}$.

Proof

By 6.10.3,

$$\begin{aligned}\overline{K} \overline{K}' &= KK' - \epsilon KK'SQ - \epsilon KSQK' + \epsilon^2 KSQK'SQ, \\ \overline{K}' \overline{K} &= K'K - \epsilon K'KSQ - \epsilon K'SQK + \epsilon^2 K'SQKSQ.\end{aligned}$$

We have $KSQK'S = KSQSK'S = 0$ by (5) and (1) of Lemma 6.10.4. Similarly, $K'SQKS = 0$. Thus, it remains to show:

$$(*) \quad KSQK' = K'SQK.$$

Let $U = KSQK'$. We will show that the matrix U has the following properties:

- (1) U has columns $t + 1, t + 2, \dots, n$ all zero.
- (2) Let $F = e_{11} + e_{22} + \dots + e_{t+1, t+1}$. Then

$$U = (FKF)S(FQF)(FK'F).$$

That is, U is the product of the top left $(t + 1) \times (t + 1)$ corners of K, S, Q, K' .

For (1) note that, by Lemma 6.10.2 (3), the nonzero entries of columns $t + 1, t + 2, \dots, n$ of K' occur only in the even rows. Hence, SK' has columns $t + 1, \dots, n$ all zero, so that by Lemma 6.10.4 (5), $U = (KSQ)(SK')$ has likewise.

To show (2), note by (1) that $U = UF$. Also $KS = FKS$ by the upper triangularity of K , so that $U = FU$. Again using triangularity, $K'F = FK'F$. Finally, $S = FSF$ is clear. Thus,

$$\begin{aligned} U &= FUF \\ &= FK(FSF)Q(FK'F) \\ &= (FKF)S(FQF)(FK'F). \end{aligned}$$

Property (2), together with its symmetric analogue for $U' = K'SQK$, shows that to prove (*), it suffices to work with the top left $(t+1) \times (t+1)$ corners of the matrices in (*). A final lemma invoking the properties in Lemma 6.10.2 (2), (4) (taking $m = t+1$ and T, R, L, L' as the $m \times m$ corners of S, Q, K, K' , respectively) completes the proof. \square

Lemma 6.10.7

Let $m \geq 3$ be an odd integer. Let T and R be the $m \times m$ matrices

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix} = \text{diag}(1, 0, 1, 0, \dots, 1),$$

$$R = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \end{bmatrix}.$$

Suppose that $L = (a_{ij})$ and $L' = (a'_{ij})$ are strictly upper triangular commuting $m \times m$ matrices whose entries satisfy

- (i) $a_{ij} = 0$ when both i and j are odd,
- (ii) $a_{ij} = a_{i+2, j+2}$ for $i < j \leq m-2$,

and the corresponding properties for the a'_{ij} . Then $LTRL' = L'TRL$.

Proof

Let V be the $m \times m$ matrix

$$V = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix}$$

and note that $V^2 = R$. From (i) and (ii) of the hypotheses, LT has the form

$$LT = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & a & 0 & b & 0 & c & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & a & 0 & b & \dots & 0 \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & \dots & \dots & 0 & 0 & a \\ 0 & 0 & 0 & \dots & \dots & 0 & 0 & 0 \end{bmatrix}$$

so that

$$LTV = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & a & 0 & b & 0 & c & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & a & 0 & b & 0 & \dots & 0 \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & \dots & \dots & 0 & a & 0 \\ 0 & 0 & 0 & \dots & \dots & 0 & 0 & 0 \end{bmatrix}.$$

Similarly,

$$VTL' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & a' & 0 & b' & 0 & c' & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & a' & 0 & b' & 0 & \dots & 0 \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & \dots & \dots & 0 & a' & 0 \\ 0 & 0 & 0 & \dots & \dots & 0 & 0 & 0 \end{bmatrix}.$$

The map that deletes the zero odd columns and the zero odd rows of the algebra of matrices having the form of LTV and VTL' is an algebra isomorphism. Under this map the images of these two matrices have the form

$$\begin{bmatrix} u & v & w & x & y & \dots \\ 0 & u & v & w & x & \dots \\ 0 & 0 & u & v & w & \dots \\ \vdots & & & & & \\ \vdots & & & & & \\ 0 & 0 & 0 & \dots & u \end{bmatrix}.$$

Such matrices commute because they are just polynomials in the $m \times m$ basic Jordan nilpotent matrix. Therefore LTV and VTL' commute whence, using Lemma 6.10.4 (5) (in the current notation this reads $TR = TRT$), we obtain

$$\begin{aligned} LTRL' &= LTRL' \\ &= LTV^2TL' \\ &= (LTV)(VTL') \\ &= (VTL')(LTV) \\ &= VT(L'L)TV \\ &= VT(LL')TV \\ &= \dots \\ &= L'TRL. \end{aligned}$$

This completes the proof of the lemma and therefore of Proposition 6.10.6. \square

Remark 6.10.8

The arguments in this section can be applied to any number of commuting matrices W, K, K', K'', \dots . Put another way, the perturbation \overline{W} in the nonhomogeneous case is k -correctable for all k . \square

We illustrate our perturbations in the nonhomogeneous case:

Example 6.10.9

Let

$$W = \left[\begin{array}{cc|cc|c|c} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline & & 0 & 0 & 1 & 0 \\ & & 0 & 0 & 0 & 0 \\ & & & & 0 & 1 \\ & & & & & 0 \end{array} \right],$$

$$K = \left[\begin{array}{cc|cc|c|c} 0 & 1 & 1 & 0 & 1 & 4 \\ 0 & 0 & 0 & 2 & 2 & 1 \\ \hline & & 0 & 1 & 1 & 1 \\ & & 0 & 0 & 0 & 2 \\ & & & & 0 & 1 \\ & & & & & 0 \end{array} \right], \quad K' = \left[\begin{array}{cc|cc|c|c} 0 & 0 & 3 & 1 & 2 & 1 \\ 0 & 0 & 0 & 3 & 0 & 2 \\ \hline & & 0 & 0 & 3 & 2 \\ & & 0 & 0 & 0 & 0 \\ & & & & 0 & 3 \\ & & & & & 0 \end{array} \right].$$

Then W is a nilpotent Weyr matrix of structure $(2, 2, 1, 1)$. One checks that K and K' commute, and from their form, we know each commutes with W . Clearing out K and K' , we get the modified matrices

$$K_1 = K - W - W^2 - 4W^3 = \left[\begin{array}{cc|cc|c|c} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ \hline & & 0 & 1 & 0 & 0 \\ & & 0 & 0 & 0 & 2 \\ & & & & 0 & 0 \\ & & & & & 0 \end{array} \right],$$

$$K'_1 = K' - 3W - 2W^2 - W^3 = \left[\begin{array}{cc|cc|c|c} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \\ \hline & & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 0 & 0 \\ & & & & 0 & 0 \\ & & & & & 0 \end{array} \right].$$

Perturbing these according to our recipe in 6.10.3, after noting $t = 4$, $E = e_{55}$, and $S = e_{11} + e_{33} + e_{55}$, gives

$$\overline{W} = W + \epsilon E = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \epsilon & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\overline{K}_1 = K_1 - \epsilon K_1 S Q$$

$$\begin{aligned} &= \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} - \\ &\quad \epsilon \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2\epsilon & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \end{aligned}$$

$$\overline{K}'_1 = K'_1 - \epsilon K'_1 S Q = K'_1. \quad \square$$

Remarks 6.10.10

- (1) If we wish to, we can also give commuting perturbations \overline{W} , \overline{K} , \overline{K}' of the original matrices by taking \overline{W} as before and letting

$$\begin{aligned} \overline{K} &= \overline{K}_1 + \overline{W} + \overline{W}^2 + 4\overline{W}^3 \\ \overline{K}' &= \overline{K}'_1 + 3\overline{W} + 2\overline{W}^2 + \overline{W}^3. \end{aligned}$$

- (2) In the above example, the Weyr structure of W involved the same number of 1's as 2's. If there are more 1's than 2's, the argument becomes

quite simple. For K and K' (in cleared form) now have a zero $t + 1$ column as well as a zero $t + 1$ row. We can take $\overline{W} = W + \epsilon E$ as before, but now take $\overline{K} = K$ and $\overline{K}' = K'$. Since $EK = 0 = KE$ and $EK' = 0 = K'E$, clearly K and K' already commute with \overline{W} . (More generally, this argument works for any Weyr structure $(n_1, \dots, n_s, 1, 1, \dots, 1)$ that has more than s 1's.) \square

We have now successfully completed the induction step in the nonhomogeneous case. Therefore we have proved the Neubauer–Sethuraman Theorem 6.9.1. However, since the argument has been spread over the last two sections, and those sections in turn have relied on earlier sections, let us sum up the process.

Proof of Theorem 6.9.1

By induction we can assume that $n \geq 2$ and that the theorem holds for matrices of size smaller than $n \times n$. Let W, K, K' be commuting $n \times n$ matrices where W is 2-regular. To establish the ASD property for these matrices, by our Reduction Principle 6.4.1 and the extended simultaneous triangularization Theorem 2.3.5, we may assume that W is a 2-regular nilpotent Weyr matrix and K, K' are strictly upper triangular. Let $\epsilon > 0$ be given. By our arguments in this and the previous section, namely the confluence of Proposition 6.9.6 (and the reduction of case (2) to case (1)) and Propositions 6.10.5, 6.10.6, we can obtain ϵ -perturbations $\overline{W}, \overline{K}, \overline{K}'$ of W, K, K' that remain commuting but where \overline{W} is a 2-regular matrix with two distinct eigenvalues 0 and ϵ (of geometric multiplicities 2 and 1, respectively). There is now a nontrivial simultaneous block diagonal splitting of $\overline{W}, \overline{K}, \overline{K}'$, courtesy of Proposition 6.4.1. On each of its blocks, \overline{W} will be (at most) 2-regular. Thus, by induction, corresponding blocks of $\overline{W}, \overline{K}, \overline{K}'$ have the ASD property and therefore so too do their parents by Proposition 6.4.1. In turn, of course, this shows that W, K, K' are ASD, as desired. \square

6.11 BOUNDS ON $\dim \mathbb{C}[A_1, \dots, A_k]$

As a corollary to Theorems 6.9.1 and 6.3.3, we obtain the following result of Neubauer and Sethuraman, whose proof involved algebraic geometry.²³

Corollary 6.11.1 (Neubauer–Sethuraman)

If A_1, A_2, A_3 are commuting $n \times n$ complex matrices and at least one is 2-regular, then $\dim \mathbb{C}[A_1, A_2, A_3] \leq n$.

23. Their proof works over any algebraically closed field.

Proof

The three matrices have the ASD property by Theorem 6.9.1, whence we have $\dim \mathbb{C}[A_1, A_2, A_3] \leq n$ by Theorem 6.3.3. \square

Example 6.3.4 shows that the ASD property can fail for more than three commuting matrices even when one of them is 2-regular. So in that case we cannot use our argument in Corollary 6.11.1 to bound the dimension of the subalgebra such matrices generate. Our techniques, however, still yield the following (sharp) upper bound. (It is not clear whether this result also follows from algebraic geometry.)

Theorem 6.11.2

Let A_1, \dots, A_k be commuting $n \times n$ matrices over the complex numbers, at least one of which is 2-regular. Then $\dim \mathbb{C}[A_1, \dots, A_k] \leq 5n/4$.

Proof

Let $\mathcal{A} = \mathbb{C}[A_1, A_2, \dots, A_k]$ with, say, A_1 a genuinely 2-regular matrix (not 1-regular, otherwise \mathcal{A} is generated by A_1 alone). By Proposition 6.4.1, we can assume that A_1 is a 2-regular nilpotent Weyr matrix and is nonzero (otherwise $n \leq 2$ and the result is easy). Since the bound to be established is independent of k , clearly there is no loss of generality in assuming that $\{A_1, \dots, A_k\}$ is a vector space basis for \mathcal{A} . (We will not assume that the other A_i are nilpotent for $i \geq 2$.) Let $W = A_1$. We consider two cases.

Case 1: W is homogeneous.

Let $r = n/2$. By the clearing out argument in the proof of Lemma 6.9.2 applied to the spanning set $\{W, A_2, \dots, A_k\}$, there exists a non-negative integer $h \leq r/2$ and a matrix $K \in \mathcal{A}$ whose W -expansion is

$$K = [D_h]W^h + [D_{h+1}]W^{h+1} + \dots + [D_{r-1}]W^{r-1},$$

and such that \mathcal{A} is spanned (as a vector space) by

$$W^0, W, W^2, \dots, W^{r-h-1}, K, KW, KW^2, \dots, KW^{r-2h-1} \quad (*)$$

and various matrices of the form

$$[D'_{r-h}]W^{r-h} + \dots + [D'_{r-1}]W^{r-1}. \quad (**)$$

The first $(*)$ group has $(r-h) + (r-2h) = 2r - 3h$ members. The second $(**)$ group clearly span a vector space of dimension at most $4h$, where the 4 comes from the number of independent choices for the D_i , and the h for the number of terms in the sum $(**)$. Therefore,

$$\dim \mathcal{A} \leq 2r - 3h + 4h = 2r + h \leq 2r + r/2 = 5r/2 = 5n/4$$

as desired.

Case 2: W is nonhomogeneous.

In this case, by Section 6.10, we can perturb A_1, A_2, \dots, A_k to commuting matrices $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k$ such that \bar{A}_1 is 2-regular with two distinct eigenvalues. (See Remark 6.10.8.) By Lemma 6.3.2, we can choose the perturbations small enough to ensure $\bar{A}_1, \dots, \bar{A}_k$ are linearly independent. Now by Proposition 6.4.1, there is a partition $n = n_1 + n_2 + \dots + n_r$ of n with $r > 1$ and a simultaneous similarity transformation of $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k$ such that

$$\bar{A}_i = \text{diag}(B_{i1}, B_{i2}, \dots, B_{ir}) \quad \text{for } i = 1, \dots, k,$$

where each B_{ij} is an $n_j \times n_j$ matrix and B_{1j} is 2-regular. For fixed j , the matrices $B_{1j}, B_{2j}, \dots, B_{kj}$ commute, whence by induction (or repeated splittings to the homogeneous case) we have $\dim \mathbb{C}[B_{1j}, B_{2j}, \dots, B_{kj}] \leq 5n_j/4$. Hence,

$$\begin{aligned} \dim \mathbb{C}[A_1, \dots, A_k] &\leq \dim \mathbb{C}[\bar{A}_1, \dots, \bar{A}_k] \\ &\leq \sum_{j=1}^r \dim \mathbb{C}[B_{1j}, \dots, B_{kj}] \\ &\leq \sum_{j=1}^r 5n_j/4 \\ &= 5n/4, \end{aligned}$$

which completes the proof. □

The following example shows that the $5n/4$ bound in Theorem 6.11.2 is sharp.

Example 6.11.3

For each positive integer n that is a multiple of 4, there is a commutative subalgebra \mathcal{A} of complex $n \times n$ matrices containing a 2-regular matrix and having $\dim \mathcal{A} = 5n/4$.

For suppose $n = 4h$. Let W be the nilpotent $n \times n$ Weyr matrix with Weyr structure $(2, 2, \dots, 2)$, that is, as a $2h \times 2h$ blocked matrix with 2×2 blocks,

$$W = \begin{bmatrix} 0 & I & & & \\ & 0 & I & & \\ & & \ddots & \ddots & \\ & & & \ddots & \\ & & & & 0 & I \\ & & & & & 0 \end{bmatrix}.$$

Let \mathcal{A} be the subalgebra of all matrices of the form

$$[D_0] + [D_1]W + [D_2]W^2 + \cdots + [D_{2h-1}]W^{2h-1}$$

where each $[D_i]$ is a block diagonal matrix with repeated 2×2 diagonal blocks D_i but with the restriction that D_0, D_1, \dots, D_{h-1} must be scalar matrices. Note that these matrices centralize W , and the product of a pair with $D_0 = D_1 = \cdots = D_{h-1} = 0$ results in zero because $W^{2h} = 0$. Thus, \mathcal{A} is commutative, contains the 2-regular matrix W , and is generated as a vector space by I, W, \dots, W^{h-1} and matrices of the form $[D_h]W^h + [D_{h+1}]W^{h+1} + \cdots + [D_{2h-1}]W^{2h-1}$. The first h powers of W contribute a dimension of h , while the matrices in the second group contribute a dimension of $4(2h - 1 - h + 1) = 4h$. Thus,

$$\dim \mathcal{A} = h + 4h = 5n/4. \quad \square$$

An interesting, but probably difficult, problem would be to find a sharp upper bound in terms of d and n on the dimension of any commutative subalgebra \mathcal{A} of $M_n(\mathbb{C})$ that contains a d -regular matrix. (The answers for $d = 1$ and $d = 2$ are, respectively, n and $5n/4$.) One conjecture might be something like

$$\dim \mathcal{A} \leq \left(\frac{1 + d^2}{2d} \right) n.$$

If the right side is an upper bound, it will be sharp by an example similar to Example 6.11.3. The proposed bound does fit the pattern for $d = 1$ and $d = 2$ (by Proposition 3.2.4 and Theorem 6.11.2). For $d = 3$, it suggests the bound of $((1 + 3^2)/6)n = 5n/3$. To establish this, it would be enough to assume \mathcal{A} contains a 3-regular nilpotent Weyr matrix W . If W has a homogeneous Weyr structure $(3, 3, \dots, 3)$, then we can indeed confirm the bound.²⁴ For let $r = n/3$ and let U_0, U_1, \dots, U_{r-1} be the leading edge subspaces of \mathcal{A} relative to W . Let h be the first index for which $\dim U_h > 1$. (If no such index exists, we have the easy bound of $r = n/3$ for $\dim \mathcal{A}$.) If $h \geq r/2$, then noting that always $\dim U_j \leq \dim M_3(F) = 9$, we have by Theorem 3.4.3

$$\begin{aligned} \dim \mathcal{A} &= \sum_{j=0}^{r-1} \dim U_j \\ &\leq h + (r - h)9 = 9r - 8h \\ &\leq 9r - 4r = 5r = 5n/3. \end{aligned}$$

24. Life has proved a lot harder in the nonhomogeneous case. The authors do not have an argument for that.

So we can assume that $h < r/2$. For $j = 1, 2, \dots, r - h - 1$, we know that U_j must centralize U_h by Proposition 3.4.4 (4). Therefore, since $\dim \mathcal{C}(Y) \leq 5$ for a nonscalar 3×3 matrix Y (this is easily checked), we have

$$\dim U_j \leq 5 \quad \text{for } j = h, h + 1, \dots, r - h - 1.$$

Now

$$\begin{aligned} \dim \mathcal{A} &= \sum_{j=0}^{r-1} \dim U_j \\ &\leq h + 5(r - 2h) + 9h \\ &= 5r = 5n/3, \end{aligned}$$

which completes the argument.

6.12 ASD FOR COMMUTING TRIPLES OF LOW ORDER MATRICES

The only known values of n for which all triples of commuting $n \times n$ matrices have ASD are $n \leq 8$. In this final section, we handle the cases $n \leq 5$, and make some comments on $n = 6, 7, 8$. By our earlier reductions, we can assume our three commuting matrices are W, K, K' , where W is a nilpotent Weyr matrix and K, K' are strictly upper triangular. By induction on the size of the matrices, it is enough to find commuting perturbations $\overline{W}, \overline{K}, \overline{K}'$ for which one of them has two distinct eigenvalues.

Cases $n = 1, 2, 3$. For these values of n , every commutative subalgebra of $M_n(F)$ is 2-generated, whence by Proposition 6.3.1 and the Motzkin–Tausky Theorem 6.8.1, ASD holds for any finite number of commuting matrices. One way to check the 2-generated claim is through Theorem 5.4.4. Or one can check it directly.

Case $n = 4$. The possible Weyr structures of W are

- (1, 1, 1, 1)
- (2, 1, 1)
- (2, 2)
- (3, 1)
- (4).

The first three are covered by our 1-regular and 2-regular ASD results (Theorem 6.9.1). The last structure is for $W = 0$, but this is covered by the Motzkin–Taussky theorem. Therefore, only structure (3, 1) remains. After clearing the (1, 4) entries of K and K' by subtracting a scalar multiple of W , we can assume

$$W = \left[\begin{array}{ccc|c} 0 & 0 & 0 & 1 \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ \hline & & & 0 \end{array} \right],$$

$$K = \left[\begin{array}{ccc|c} 0 & a & b & 0 \\ & 0 & c & d \\ & & 0 & e \\ \hline & & & 0 \end{array} \right], \quad K' = \left[\begin{array}{ccc|c} 0 & a' & b' & 0 \\ & 0 & c' & d' \\ & & 0 & e' \\ \hline & & & 0 \end{array} \right].$$

We can assume that no linear combination of W, K, K' is 2-regular, that is, has nullity smaller than 3 (k -regular for a nilpotent matrix means its null space, which is the eigenspace of its single eigenvalue 0, has dimension at most k). Otherwise we could replace K or K' by a 2-regular nilpotent matrix and be back to an earlier covered case. Thus, we can assume:

every matrix in the linear span of W, K, K' has rank at most 1. (*)

It follows that either (1) the first rows of K and K' are zero, or (2) their last columns are zero. When (1) holds, we perturb W to

$$\overline{W} = W + \epsilon e_{11} = \left[\begin{array}{ccc|c} \epsilon & 0 & 0 & 1 \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ \hline & & & 0 \end{array} \right],$$

while for (2) we use

$$\overline{W} = W + \epsilon e_{44} = \left[\begin{array}{ccc|c} 0 & 0 & 0 & 1 \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ \hline & & & \epsilon \end{array} \right].$$

In each situation, K and K' already commute with \overline{W} because e_{11} (respectively e_{44}) annihilates K and K' on both sides. So K and K' don't require a matching perturbation. This completes the argument for structure $(3, 1)$.

Remark

We know from Example 6.3.4 that these arguments can't work for more than three commuting 4×4 matrices.

Case n = 5. The possible Weyr structures of W are

$$(1, 1, 1, 1, 1)$$

$$(2, 1, 1, 1)$$

$$(2, 2, 1)$$

$$(3, 1, 1)$$

$$(3, 2)$$

$$(4, 1)$$

$$(5).$$

The first three are covered by our earlier 1-regular and 2-regular general results. The last is for $W = 0$, which is covered by the Motzkin–Tausky theorem for commuting pairs. That leaves us the three subcases $(3, 1, 1)$, $(3, 2)$, and $(4, 1)$, of which only the middle one presents any real challenge.

Subcase: structure $(3, 1, 1)$. This is handled in the exactly the same way as a nonhomogeneous structure $(2, \dots, 2, 1, \dots, 1)$ with more 1's than 2's. Namely, we use W to clear the $(1, 4)$ entries of K and K' , and then observe that the new K and K' have a zero fourth row and column. We take $\overline{W} = W + \epsilon e_{44}$, $\overline{K} = K$, and $\overline{K}' = K'$. See Remark 6.10.10 (2).

Subcase: structure $(3, 2)$. Here the three bears look like

$$W = \left[\begin{array}{ccc|cc} 0 & 0 & 0 & 1 & 0 \\ & 0 & 0 & 0 & 1 \\ & & 0 & 0 & 0 \\ \hline & & & 0 & 0 \\ & & & 0 & 0 \end{array} \right],$$

$$K = \left[\begin{array}{ccc|cc} 0 & a & b & d & e \\ & 0 & c & f & g \\ & & 0 & h & i \\ \hline & & & 0 & a \\ & & & & 0 \end{array} \right], \quad K' = \left[\begin{array}{ccc|cc} 0 & a' & b' & d' & e' \\ & 0 & c' & f' & g' \\ & & 0 & h' & i' \\ \hline & & & 0 & a' \\ & & & & 0 \end{array} \right].$$

We can also assume that, for any X in the linear span of W, K, K' , we have

- (i) $\text{rank}(X) \leq 2$ and
- (ii) the nilpotency index of X is at most 2.

Indeed, if (i) fails, then nullity $X \leq 2$, which makes X a 2-regular matrix, while if (ii) fails then the Weyr structure of X is (n_1, n_2, \dots, n_r) for some $r \geq 3$. In either case, we can replace K or K' by X to bring us back to a subcase already covered. For instance, if $X = 2W + 5K - K'$ had index 3, we could consider the commuting triple W, K, X of nilpotents where the Weyr structure of X is either $(2, 2, 1)$ or $(3, 1, 1)$. Note that the new matrices generate the same subalgebra as the old ones, so they share or fail ASD together.

Suppose $a \neq 0$. Then, by (i), $c = f = h = 0$. By subtracting scalar multiples of W from K and K' , we can make $d = d' = 0$. By subtracting a scalar multiple of K from K' we can make $a' = 0$. Thus, K and K' look like

$$K = \left[\begin{array}{ccc|cc} 0 & a & b & 0 & e \\ & 0 & 0 & 0 & g \\ & & 0 & 0 & i \\ \hline & & & 0 & a \\ & & & & 0 \end{array} \right], \quad K' = \left[\begin{array}{ccc|cc} 0 & 0 & b' & 0 & e' \\ & 0 & c' & f' & g' \\ & & 0 & h' & i' \\ \hline & & & 0 & 0 \\ & & & & 0 \end{array} \right].$$

Let $E = e_{24}$ and observe that $EK' = 0 = K'E$ and $EW = 0 = WE$. Thus, the perturbation

$$\bar{K} = K + \epsilon E = \left[\begin{array}{ccc|cc} 0 & a & b & 0 & e \\ & 0 & 0 & \epsilon & g \\ & & 0 & 0 & i \\ \hline & & & 0 & a \\ & & & & 0 \end{array} \right]$$

preserves commutativity with W and K' . But now \bar{K} is a nilpotent of rank 3, hence 2-regular. We are back to an earlier subcase. Therefore, we can assume $a = 0$ and, by symmetry, that $a' = 0$.

Claim: Without loss of generality, we can assume

$$a = b = c = 0,$$

$$a' = b' = c' = 0.$$

We know we can assume $a = a' = 0$. If $b \neq 0$, then by (ii), $h = i = 0$ otherwise $K^2 \neq 0$; and $h' = i' = 0$ otherwise $(\alpha K + K')^2 \neq 0$ for a suitable scalar α . The same conclusion is reached if one of c, b', c' is nonzero. Thus, were the claim to fail, we would have to have $h = i = 0$ and $h' = i' = 0$. Now for a cute observation. We all know that transposing in the main NW-SE diagonal is an algebra anti-automorphism (a linear isomorphism that reverses products). The same is true if we transpose in the other NE-SW diagonal! This is a little-known fact, but easily checked.²⁵ Call the second transpose across the NE-SW diagonal τ , and note that τ preserves the norm $\| \cdot \|$ and maps diagonal matrices to diagonal matrices. Thus, W, K, K' have ASD if and only if $\tau(W), \tau(K), \tau(K')$ have ASD. Note that τ fixes W .²⁶ Finally, observe that when we apply τ to K , a stays put, h interchanges with c , and i interchanges with b . Similarly, in $\tau(K')$, a' stays put, h' interchanges with c' , and i' interchanges with b' . Our claim is established.

Henceforth, we make the assumption in the claim. Before introducing our perturbations of W, K, K' , we need to further modify K and K' . First we can perturb the top right 2×2 corner Z of K to a diagonalizable matrix (see Example 6.2.2). This doesn't alter commutativity with W or K' because the matrices still annihilate each other. Choose $P \in GL_2(\mathbb{C})$ such that $P^{-1}ZP$ is diagonal. Then conjugating K by $\text{diag}(P, 1, P)$ makes the top right 2×2 corner of K a diagonal matrix. Apply the conjugation also to W (it doesn't change) and K' . Finally, make the $(2, 5)$ entries of the new K and K' zero by subtracting suitable scalar multiples of W . Our remodeling is complete. To sum up, we can assume

$$K = \left[\begin{array}{ccc|cc} 0 & 0 & 0 & d & 0 \\ & 0 & 0 & 0 & 0 \\ & & 0 & h & i \\ \hline & & & 0 & 0 \\ & & & & 0 \end{array} \right], \quad K' = \left[\begin{array}{ccc|cc} 0 & 0 & 0 & d' & e' \\ & 0 & 0 & f' & 0 \\ & & 0 & h' & i' \\ \hline & & & 0 & 0 \\ & & & & 0 \end{array} \right].$$

25. Eugene Spiegel has a lovely 2005 article on this and related mappings. Jacobson on p. 243 of his *Basic Algebra II* also mentions a more general result as an exercise.

26. It is seemingly contradictory that a nonzero nilpotent matrix could be “symmetric” with respect to a transpose! It can't happen with the usual transpose over say the reals, because symmetrics are diagonalizable.

Let

$$\overline{W} = \left[\begin{array}{ccc|cc} 0 & 0 & 0 & 1 & 0 \\ & \epsilon & 0 & 0 & 1 \\ & & 0 & 0 & 0 \\ \hline & & & 0 & 0 \\ & & & 0 & 0 \end{array} \right],$$

$$\overline{K} = \left[\begin{array}{ccc|cc} 0 & 0 & 0 & d & 0 \\ & 0 & 0 & 0 & 0 \\ & & \epsilon_1 & h & i \\ \hline & & & 0 & 0 \\ & & & & 0 \end{array} \right], \quad \overline{K}' = \left[\begin{array}{ccc|cc} 0 & 0 & 0 & d' & e' \\ & 0 & 0 & f' & 0 \\ & & 0 & h' & i' \\ \hline & & & 0 & 0 \\ & & & -\epsilon f' & 0 \end{array} \right],$$

where ϵ_1 is an order of epsilon term to be determined. One checks that \overline{W} commutes with \overline{K} and \overline{K}' . (The slick way to do this is to adopt the argument used in 6.9.4 for the homogeneous structure $(2, 2)$.) The condition for \overline{K} and \overline{K}' to commute is that

$$\epsilon_1 h' = \epsilon i f' \text{ and } \epsilon_1 i' = 0. \quad (*)$$

If $i = 0$ we can meet $(*)$ by taking $\epsilon_1 = 0$. Suppose $i \neq 0$. By subtracting a multiple of K from K' , we can assume $i' = 0$. Thus, if $h' \neq 0$ we can meet $(*)$ with $\epsilon_1 = \epsilon i f' / h'$. The final remaining situation is when $h' = i' = 0$, but this can be handled by a different and simpler perturbation. Namely we observe that now $E = e_{33}$ annihilates W and K' on both sides, so we can take

$$\overline{W} = W, \quad \overline{K} = K + \epsilon E, \quad \overline{K}' = K'.$$

With both sets of perturbations, we have achieved commuting triples such that one matrix has 0 and ϵ as eigenvalues; \overline{W} in the first instance, \overline{K} in the second. We have now finished the subcase of W having structure $(3, 2)$.

Subcase: structure $(4, 1)$. This is handled almost identically to the earlier structure $(3, 1)$ for $n = 4$. Because it is now the only case outstanding, we can assume that no linear combination X of W , K and K' has rank bigger than 1 (otherwise X is 3-regular or less). Use W to clear the $(1, 5)$ entries of K and K' . By rank considerations, we see that either the first rows of K and K' are both

zero, or their last columns are zero. In the first case, perturb W to $W + \epsilon e_{11}$, leaving K and K' unchanged. In the second, perturb W to $W + \epsilon e_{55}$.

This completes the case $n = 5$. □

The history of the low order cases where triples (A_1, A_2, A_3) of commuting $n \times n$ complex matrices have been shown to possess ASD is briefly as follows. Clearly, $n = 1, 2$ are easy and have been known since ASD was first formulated (probably the 1950s). The case $n = 3$ was handled by Guralnick in 1992, and $n = 4$ by Guralnick and Sethuraman in 2000. To be fair, in both cases, a much more general result was established in terms of algebraic varieties (see Chapter 7). Holbrook and Omladić did $n = 5$ in 2001, Omladić $n = 6$ in 2004, Han $n = 7$ in 2005, and Šivic $n = 8$ in 2008. Again, these latter three results are corollaries to more general results, often over an algebraically closed field of characteristic zero and with approximation relative to the Zariski topology. On the other hand, it is hard to see how their arguments would be any shorter if only the complex field and the Euclidean metric were used. All use a case-by-case analysis in terms of the Jordan structure of an appropriate nilpotent matrix. And the details can be quite involved. For instance, Han's $n = 7$ runs to some 70 pages, although Šivic's arguments have since reduced this.

Whether it be wishful thinking or not (we can't face $n = 9$ running to over 100 pages!), a general feeling among researchers is that ASD must fail pretty soon after $n = 8$. But who knows? In any event, as earlier commented, we certainly don't have to go beyond $n = 28$. Some of our readers, those possessing the boldness of a bald eagle,²⁷ will undoubtedly take up the challenge of finding the exact cutoff.²⁸

The ASD question is interesting and challenging—a good test of a canonical form. We feel that with ASD questions, the Weyr form is a more promising tool than its Jordan counterpart. We hope that this chapter has strengthened our position, while pointing the way towards the next advance.

BIOGRAPHICAL NOTES ON MOTZKIN AND TAUSSKY

Theodore Samuel Motzkin was born on March 26, 1908, in Berlin. His mathematical ability became apparent at an early age and he started tertiary

27. The bald eagle is the national bird and symbol of the United States, but is found throughout North America. A large bird with a wingspan of up to 2.5 meters, the bald eagle soars on thermal convection currents. Its dive speed can reach 160 km per hour. Fish are its standard prey, but also rabbits, raccoons, ducks, even deer fawn, can be on the menu.

28. Along with the kookaburra and kea, this completes the troika of birds chosen to match the countries of residence of the three authors (and their respective personalities).

studies at the age of 15. As was customary in Germany in those days, he spent time at various universities, including Göttingen, Paris, and Berlin. He finished his diploma thesis on algebraic structures under Schur in Berlin, and then went to Basel for his doctorate under Mostrowski, working on linear inequalities, completing in 1934. Linear programming, power series, geometric problems, and graph theory became main themes of his research, but his mathematical interests were very broad. In 1935 he was appointed to the Hebrew University in Jerusalem, and during the war years he worked there as a cryptographer for the British government. He emigrated to the United States in 1948. In 1949, the *Bulletin of the American Mathematical Society* published his paper “The Euclidean Algorithm” in which he cleverly exhibited classes of principal ideal domains that are not Euclidean domains, including the oft-quoted but seldom-detailed $\mathbb{Z}[(1 + \sqrt{-19})/2]$. He joined the University of California, Los Angeles, in 1950, becoming professor in 1960. He died on December 15, 1970, in Los Angeles.

Olga Taussky was born on August 30, 1906, in what is now known as Olomouc, in the Czech Republic. In 1916 the Taussky family moved to Linz in Austria and later Olga entered the University of Vienna with mathematics as her main subject. Her doctorate was on algebraic number fields, completed under Philipp Furtwängler in 1930, just as class field theory appeared on the scene. In 1931 Courant appointed her as an assistant at Göttingen where she helped to edit Hilbert’s complete works on number theory and assisted Artin and Noether with their class field theory notes and lectures. After brief spells in the United States (at Bryn Mawr, with Emmy Noether) and Cambridge, in 1937 she obtained a teaching position at a college in London and shortly after married fellow mathematician John Todd. With a leave of absence from her teaching, in 1943 to 1946 she worked on aerodynamics problems at Britain’s National Physical Laboratory and here she “realised the beauty of research on differential equations” and “learned a lot of matrix theory.” In 1957 both she and her husband joined the staff at the California Institute of Technology. She wrote about 300 papers, mostly in matrix theory, group theory, and algebraic number theory. She died in Pasadena, California, on October 7, 1995.

Algebraic Varieties

Algebraic varieties are the stuff of algebraic geometry. But what has algebraic geometry got to do with our linear algebra problems? Quite a lot, as it turns out, because the ASD (approximate simultaneous diagonalization) question for k commuting $n \times n$ matrices over \mathbb{C} , which we studied in Chapter 6, is equivalent to the irreducibility of a certain affine variety of matrices over \mathbb{C} . Not only that, in certain cases it is easier to establish that irreducibility, or lack thereof, than it is to establish ASD (or its failure) directly. For instance, the only proof that the authors are aware of that shows commuting triples of $n \times n$ complex matrices fail the ASD property in general for all $n \geq 29$ is through Guralnick's use of algebraic geometry. Moreover, the proofs are most elegant.

We aim to use the traditional license of authors in their final chapter to take a branch off the main road and give a largely self-contained account of the algebraic geometry connection to our linear algebra problems. No prior knowledge of algebraic geometry is required. This is perhaps an ambitious undertaking on the authors' part, and does require a higher level of sophistication of the reader than in earlier chapters. But an understanding of the material is well within the grasp of a good graduate student who knows the basics of elementary commutative algebra and elementary topology. And we believe the rewards are great.

Too few mathematicians are aware of the power and beauty of algebraic geometry. Most are aware that (somehow) algebraic geometry has played a major role in number theory, such as in Wiles's solution¹ of Fermat's Last Theorem, but often folk are completely unaware of how algebraic geometry impacts their own speciality, be that in linear algebra, for instance. Algebraic geometry is thought of as being a very difficult subject to understand. And in deep applications that is true for many of us. But the applications we have in mind require little beyond elementary algebraic geometry. Still, to develop this from scratch does require some work. We have taken pains to explain things simply (we hope!), with a general audience in mind. To the expert, some of our arguments may appear a little labored.

In Sections 7.1 to 7.4, we present some of the basics of algebraic geometry: affine varieties, polynomial maps, the Zariski topology, Hilbert's basis theorem, Hilbert's nullstellensatz, Noether's normalization theorem, and irreducible varieties.

Section 7.5 establishes the equivalence of the ASD property for k commuting $n \times n$ complex matrices with the irreducibility of the variety $\mathcal{C}(k, n)$ of k -tuples of commuting $n \times n$ complex matrices. We examine the implications of this later in the chapter.

In 1955, Motzkin and Taussky showed that $\mathcal{C}(2, n)$ is irreducible over an algebraically closed field. In Section 7.6, we present a short proof of this due to Guralnick. We also include an argument, again due to Guralnick, showing how Gerstenhaber's theorem (studied in Chapter 5) can be quickly deduced from the Motzkin–Taussky theorem. This is a lovely application of algebraic geometry.

Irreducibility of $\mathcal{C}(k, n)$ over a general algebraically closed field F is completely understood except when $k = 3$: it holds universally for $k = 1, 2$, and fails for $k \geq 4$ when $n \geq 4$. On the other hand, irreducibility of $\mathcal{C}(3, n)$ has still not been completely settled. As of 2010, $\mathcal{C}(3, n)$ is known to be irreducible for $n \leq 8$ when F has characteristic zero, and is known to be reducible for $n \geq 29$ in arbitrary characteristics. In Section 7.9, we treat the case $n \geq 29$ using a construction of Guralnick (and refined by Holbrook and Omladič). Here, our use of the Weyr form simplifies some earlier arguments and points to possible further extensions. The concept of dimension of a variety plays a critical role in these arguments. In anticipation of this, we present in Section 7.8 the basic properties of dimension, following a brief discussion of co-ordinate rings in Section 7.7.

1. A web search reveals a multitude of popular articles, books, videos, films and documentaries on Andrew Wiles's magnificent achievement in 1995, perhaps the greatest triumph of twentieth-century pure mathematics.

The material in Section 7.10 also concerns irreducibility of $\mathcal{C}(3, n)$ and other varieties of matrices but is somewhat more specialized. Some readers may prefer to skip this section, or to skim it just to see the role played by the Weyr form—that, after all, is our central theme.

The final Section 7.11 outlines a proof of a “Denseness Theorem,” used in Section 7.5, that relates Zariski denseness and classical Euclidean denseness in the case of irreducible complex affine varieties.

7.1 AFFINE VARIETIES AND POLYNOMIAL MAPS

Algebraic geometry can be described as the study of the (common) zeros of a set of polynomials, but perhaps more accurately as the study of polynomial maps and the spaces on which they act. For the present, F can be an arbitrary field. For a positive integer n , if we ignore the vector space structure of the set F^n of all n -tuples (a_1, a_2, \dots, a_n) over F , and regard its elements as just points, then this set is referred to as **affine n -space** (over F), and is denoted by \mathbb{A}^n . Note that the origin is not singled out in affine n -space. For our purposes, we define an **affine variety** V to be any subset of \mathbb{A}^n that consists of the set of all common zeros of some collection S of polynomials in $F[x_1, x_2, \dots, x_n]$:

$$V = V(S) = \{a \in \mathbb{A}^n : f(a) = 0 \text{ for all } f \in S\}.$$

We refer to $V(S)$ as the **variety determined by S** . There is no restriction placed on the set S . However, later we will see that S can be assumed to be finite.

Remark 7.1.1

One has to be careful, when reading books and articles that use algebraic geometry, to check what definition of “variety” is being used. It varies considerably, depending on the level of sophistication. Ours would be the most simple-minded. (Some authors would call our affine varieties “affine algebraic subsets.”) Indeed, prior to the 1960s, an “affine variety” usually had irreducibility (discussed in Section 7.4) built into the definition. That is rarely the case nowadays. The definition of an affine variety for serious algebraic geometers is an abstract one that does not assume a preferred embedding in \mathbb{A}^n . They also often prefer to work with “projective varieties” or “quasi-projective varieties.” \square

Of particular interest is when $S = \{f\}$ consists of a single polynomial $f = f(x_1, x_2, \dots, x_n)$, in which case

$$V(S) = V(f) = \{(a_1, a_2, \dots, a_n) \in \mathbb{A}^n : f(a_1, a_2, \dots, a_n) = 0\}$$

and is called the **hypersurface** determined by f . The two extremes occur when f is a constant polynomial, namely \mathbb{A}^n when f is zero, and the empty set \emptyset when f is a nonzero constant. The reader will be familiar with many examples of hypersurfaces in \mathbb{A}^2 over $F = \mathbb{R}$, such as ellipses, parabolas, and hyperbolas.² Hypersurfaces here are plane curves. The three plane cubics in the following example are purloined from K. Hulek's book *Elementary Algebraic Geometry*. We refer the interested reader to pp. 5–9 of that text for the full details. (We won't refer back to these examples but mention them simply for cultural reasons.)

Example 7.1.2

Consider the real cubic curves

$$C_1 : y^2 = x^3 + x^2,$$

$$C_2 : y^2 = x^3,$$

$$C_3 : y^2 = x(x-1)(x-2),$$

whose graphs are depicted in the three figures on the next page. (Of course, these are the graphs of the real affine varieties $V(y^2 - x^3 - x^2)$, $V(y^2 - x^3)$, and $V(y^2 - x(x-1)(x-2))$, respectively.) The curves C_1 and C_2 admit rational (in fact, polynomial) parameterizations, namely

$$\phi_1 : \mathbb{R} \longrightarrow \mathbb{R}^2, \quad t \longmapsto (t^2 - 1, t^3 - t)$$

$$\phi_2 : \mathbb{R} \longrightarrow \mathbb{R}^2, \quad t \longmapsto (t^2, t^3).$$

But it is not possible to rationally parameterize C_3 (even over \mathbb{C}). In fact, there is no nonconstant rational map $(f, g) : \mathbb{R} \rightarrow C_3$, $t \mapsto (f(t), g(t))$, where $f, g \in \mathbb{R}(t)$. The double point $(0, 0)$ of C_1 (corresponding to $\phi_1(-1) = \phi_1(1)$), and the cusp $(0, 0)$ of C_2 , are examples of “singular” points, and the other points are “smooth” or “regular.” Although the concepts of singular points and smooth points have an indispensable role in nonelementary algebraic geometry, our elementary treatment manages to avoid them. \square

2. Over the real field $F = \mathbb{R}$, all affine varieties are hypersurfaces because $V(f_1, \dots, f_k) = V(f_1^2 + \dots + f_k^2)$ for any $f_1, \dots, f_k \in F[x_1, \dots, x_n]$. This is certainly not true for affine varieties over an algebraically closed field, because then a hypersurface determined by a nonconstant polynomial must have “algebraic geometry dimension” $n - 1$, whereas subvarieties in general can take any dimension values from $\{0, 1, 2, \dots, n\}$. All this is covered in Section 7.8.

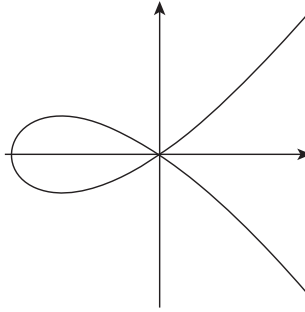


Figure 7.1 $C_1 : y^2 = x^3 + x^2$

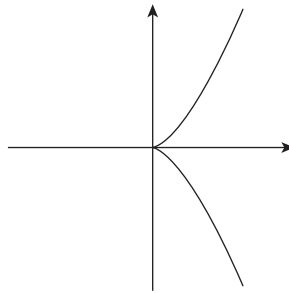


Figure 7.2 $C_2 : y^2 = x^3$

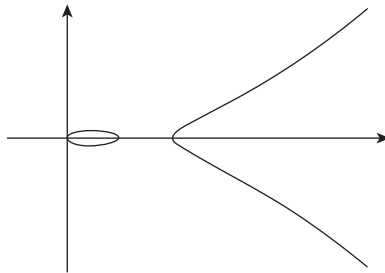


Figure 7.3 $C_3 : y^2 = x(x-1)(x-2)$

Example 7.1.3

Vector subspaces V of F^n , and their translates, are affine varieties. To see this, note that V is the solution space of some system of linear equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0, \end{aligned}$$

whence V is the variety determined by the m degree one polynomials

$$a_{11}x_1 + \cdots + a_{1n}x_n, \dots, a_{m1}x_1 + \cdots + a_{mn}x_n.$$

If $W = \mathbf{b} + V$ is a translate of V , where $\mathbf{b} = (b_1, \dots, b_n)$, then W is the variety determined by the polynomials

$$a_{11}(x_1 - b_1) + \cdots + a_{1n}(x_n - b_n), \dots, a_{m1}(x_1 - b_1) + \cdots + a_{mn}(x_n - b_n). \quad \square$$

We assume our reader is comfortable with the basics of polynomials in several variables and over any field F . (An informal “aside” in Section 7.7 should clarify some aspects for the nonconfident, and perhaps those readers should break at this point and read the material at the beginning of that section.) Let $V \subseteq \mathbb{A}^n$ and $W \subseteq \mathbb{A}^m$ be affine varieties. A **polynomial map** $f : V \rightarrow W$ is a function for which there are polynomials $p_1, p_2, \dots, p_m \in F[x_1, x_2, \dots, x_n]$ such that

$$f(a_1, a_2, \dots, a_n) = (p_1(a_1, a_2, \dots, a_n), \dots, p_m(a_1, a_2, \dots, a_n)) \in W$$

for all $(a_1, a_2, \dots, a_n) \in V$. In general, the p_i are not uniquely determined by f . A polynomial map $f : V \rightarrow W$ is an **isomorphism of varieties** if there exists a polynomial map $g : W \rightarrow V$ such that $f \circ g = 1_W$ and $g \circ f = 1_V$. (Here, $f \circ g$ is the composition of the functions f and g , with g acting first, while 1_W is the identity function on W .) Of course, we then say that the varieties V and W are **isomorphic** if there is an isomorphism $f : V \rightarrow W$ and in this case write $V \cong W$.

Example 7.1.4

- (i) As a simple example, if $V = \mathbb{A}^1$ and W is the parabola $\{(x, y) : y - x^2 = 0\}$ in \mathbb{A}^2 , the polynomial map $f : \mathbb{A}^1 \rightarrow W$, $x \mapsto (x, x^2)$ is an isomorphism with inverse map $g : W \rightarrow \mathbb{A}^1$, $(x, y) \mapsto x$.
- (ii) A polynomial map $f : V \rightarrow W$ that is a bijection need not be an isomorphism of varieties. For let $V = \mathbb{A}^1$, $W = \{(x, y) : y^2 = x^3\}$, and let

$$f : V \rightarrow W, \quad x \mapsto (x^2, x^3).$$

Then f is a bijection but the inverse map $g : W \rightarrow V$

$$g(x, y) = \begin{cases} y/x & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

is not a polynomial map. □

The varieties of most interest to us in connection with linear algebra are, naturally enough, connected with matrices. Here are some examples, the first of which is the most important.

Example 7.1.5

Let k and n be fixed positive integers. Let $\mathcal{C}(k, n)$ be the set of all k -tuples of commuting $n \times n$ matrices over F :

$$\mathcal{C}(k, n) = \{(A_1, A_2, \dots, A_k) : \text{each } A_i \in M_n(F) \text{ and } A_i A_j = A_j A_i \text{ for all } i, j\}.$$

Then $\mathcal{C}(k, n)$ can be regarded as an affine variety in affine space \mathbb{A}^{kn^2} . Why is this? First, by running through the entries of an $n \times n$ matrix A in some fixed order, say across the rows starting with the first, we can view A as a member of \mathbb{A}^{n^2} . A k -tuple of $n \times n$ matrices then sits inside $\mathbb{A}^{n^2 + \dots + n^2} = \mathbb{A}^{kn^2}$. Thus, we have identified $\mathcal{C}(k, n)$ with a certain subset V of \mathbb{A}^{kn^2} . We now want a set S of polynomials in kn^2 variables such that V is the locus of zeros of S . But the commutativity of A_i and A_j is equivalent to the commutator condition

$$A_i A_j - A_j A_i = 0,$$

which in turn can be expressed in terms of n^2 polynomial equations in $2n^2$ variables on the entries of the two matrices (all the polynomials are homogeneous of degree 2). Therefore, the condition that (A_1, A_2, \dots, A_k) be a k -tuple of commuting matrices is equivalent to the corresponding element of V vanishing at a certain set S of $n^2 k(k-1)/2$ polynomials in kn^2 variables (in fact, each polynomial is homogeneous of degree 2 and with $2n$ terms). For instance, consider $\mathcal{C}(2, 2)$. We can make the identification

$$\mathcal{C}(2, 2) \longrightarrow V \subseteq \mathbb{A}^8, \quad \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} e & f \\ g & h \end{bmatrix} \right) \mapsto (a, b, c, d, e, f, g, h)$$

where V is the affine variety determined by the following four polynomials in 8 variables:

$$\begin{aligned} p_1(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) &= x_1 x_5 + x_2 x_7 - x_5 x_1 - x_6 x_3, \\ p_2(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) &= x_1 x_6 + x_2 x_8 - x_5 x_2 - x_6 x_4, \\ p_3(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) &= x_3 x_5 + x_4 x_7 - x_7 x_1 - x_8 x_3, \\ p_4(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) &= x_3 x_6 + x_4 x_8 - x_7 x_2 - x_8 x_4. \end{aligned}$$

Often with this type of variety, it is not necessary to know explicitly the polynomials that determine the variety. It is enough to realize that they exist. \square

Example 7.1.6

The set of all $n \times n$ idempotent matrices forms an affine variety in \mathbb{A}^{n^2} because the idempotent condition $A^2 = A$ can be expressed by n^2 degree two polynomial equations in the entries of A . For example, the variety of all 2×2 idempotent matrices

$$A = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$$

is determined by the four polynomials

$$\begin{aligned} p_1(x_1, x_2, x_3, x_4) &= x_1^2 + x_2x_3 - x_1, \\ p_2(x_1, x_2, x_3, x_4) &= x_1x_2 + x_2x_4 - x_2, \\ p_3(x_1, x_2, x_3, x_4) &= x_3x_1 + x_4x_3 - x_3, \\ p_4(x_1, x_2, x_3, x_4) &= x_3x_2 + x_4^2 - x_4. \end{aligned}$$

□

Example 7.1.7

The set of all $n \times n$ nilpotent matrices likewise forms an affine variety in \mathbb{A}^{n^2} , because the nilpotent condition is $A^n = 0$ (by the Cayley–Hamilton theorem and Proposition 1.1.1). In turn this can be expressed as n^2 polynomial equations of degree n in the entries of A . For example, the variety of all 3×3 nilpotent matrices

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix}$$

is determined by 9 polynomials in $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9$. For instance, one of these polynomials

$$p_1(x_1, \dots, x_9) = x_1^3 + 2x_1x_2x_4 + 2x_1x_3x_7 + x_2x_4x_5 + x_2x_6x_7 + x_3x_4x_8 + x_3x_7x_9$$

comes from equating the $(1, 1)$ entry of A^3 to 0.

□

Example 7.1.8

Left (or right) multiplication by a fixed $n \times n$ matrix T affords a very natural example of a polynomial map $p : M_n(F) \rightarrow M_n(F)$, $A \mapsto TA$ (again identifying $M_n(F)$ with \mathbb{A}^{n^2}). For instance, let

$$T = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}.$$

Then

$$T \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = \begin{bmatrix} p_1(a_1, a_2, a_3, a_4) & p_2(a_1, a_2, a_3, a_4) \\ p_3(a_1, a_2, a_3, a_4) & p_4(a_1, a_2, a_3, a_4) \end{bmatrix}$$

where

$$\begin{aligned} p_1(x_1, x_2, x_3, x_4) &= x_1 + 2x_3, \\ p_2(x_1, x_2, x_3, x_4) &= x_2 + 2x_4, \\ p_3(x_1, x_2, x_3, x_4) &= -x_1, \\ p_4(x_1, x_2, x_3, x_4) &= -x_2. \end{aligned}$$

□

Example 7.1.9

For a positive integer n , let $SL_n(F)$ be the *special linear group* consisting of all $n \times n$ matrices over F of determinant 1. Then $SL_n(F)$ can be regarded as an affine variety in affine space \mathbb{A}^{n^2} . In fact, using our earlier identification of $n \times n$ matrices with points in \mathbb{A}^{n^2} , we see that $SL_n(F)$ is the hypersurface determined by the polynomial in n^2 variables³

$$f(x_{11}, x_{12}, \dots, x_{nn}) = \det \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix} - 1.$$

Note that f has $n!$ nonconstant terms, each of degree n . For instance, $SL_3(F)$ is the variety determined by the polynomial

$$\begin{aligned} f(x_{11}, x_{12}, \dots, x_{33}) &= x_{11}x_{22}x_{33} + x_{12}x_{23}x_{31} + x_{13}x_{21}x_{32} \\ &\quad - x_{13}x_{22}x_{31} - x_{11}x_{23}x_{32} - x_{12}x_{21}x_{33} - 1. \end{aligned}$$

□

Identifying the set $M_n(F)$ of all $n \times n$ matrices with \mathbb{A}^{n^2} , we see that the determinant map $\det : M_n(F) \rightarrow \mathbb{A}$ provides a good example of a polynomial map of affine varieties. So does the trace map.

The following observation is a valuable tool for confirming that particular sets of matrices are indeed varieties:

3. In this type of setting, we use matrix indexing for the n^2 variables.

Proposition 7.1.10

Let $p(x_1, x_2, \dots, x_n)$ be a symmetric polynomial^t in the x_i with coefficients from an algebraically closed field F . Then the mapping

$$M_n(F) \longrightarrow F, A \longmapsto p(\lambda_1, \lambda_2, \dots, \lambda_n),$$

where the λ_i are the eigenvalues of A , is a polynomial function on $M_n(F)$.

Proof

Let $A = (a_{ij}) \in M_n(F)$, $f(x) = \det(xI - A)$ be the characteristic polynomial of A , and $\lambda_1, \lambda_2, \dots, \lambda_n$ be the not necessarily distinct eigenvalues of A . Evaluating the determinant gives

$$(1) \quad f(x) = x^n + g_1 x^{n-1} + \dots + g_{n-1} x + g_n$$

for some polynomials g_i in the matrix entries $a_{11}, a_{12}, \dots, a_{nn}$. On the other hand,

$$f(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$$

and expanding this gives

$$(2) \quad f(x) = x^n - s_1 x^{n-1} + s_2 x^{n-2} + \dots + (-1)^n s_n$$

where the s_i are the elementary symmetric polynomials in $\lambda_1, \lambda_2, \dots, \lambda_n$:

$$\begin{aligned} s_1 &= \lambda_1 + \lambda_2 + \dots + \lambda_n \\ s_2 &= \sum_{i < j} \lambda_i \lambda_j \\ &\vdots \\ s_n &= \lambda_1 \lambda_2 \cdots \lambda_n. \end{aligned}$$

Equating the coefficients in (1) and (2) shows that the elementary symmetric polynomials s_1, s_2, \dots, s_n in the *eigenvalues* of A actually agree with certain polynomials in the *entries* of A . Consequently, the same is true of a general symmetric polynomial $p(\lambda_1, \dots, \lambda_n)$ because such a p takes the form

$$p(\lambda_1, \dots, \lambda_n) = q(s_1(\lambda_1, \dots, \lambda_n), \dots, s_n(\lambda_1, \dots, \lambda_n))$$

for some polynomial $q \in F[x_1, \dots, x_n]$ (see Jacobson's *Basic Algebra I*, p. 135). \square

We illustrate the proposition with the following:

4. A polynomial $p(x_1, x_2, \dots, x_n)$ is *symmetric* if $p(x_1, x_2, \dots, x_n) = p(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)})$ for all permutations σ of $\{1, 2, \dots, n\}$.

Example 7.1.11

(i) The symmetric polynomial

$$p(\lambda_1, \lambda_2, \dots, \lambda_n) = \lambda_1 + \lambda_2 + \dots + \lambda_n$$

in the eigenvalues λ_i of an $n \times n$ matrix $A = (a_{ij})$ can be expressed as the trace polynomial

$$\text{tr } A = a_{11} + a_{22} + \dots + a_{nn}$$

in the entries of A . Likewise, the symmetric polynomial

$$p(\lambda_1, \lambda_2, \dots, \lambda_n) = \lambda_1 \lambda_2 \dots \lambda_n$$

in the eigenvalues λ_i of A can be expressed as the determinant polynomial $\det A$ in the entries of A .

(ii) The following polynomial

$$p(\lambda_1, \lambda_2, \dots, \lambda_n) = \prod_{i \neq j} (\lambda_i - \lambda_j)$$

is also clearly symmetric in the eigenvalues of an $n \times n$ matrix A . While a corresponding polynomial expression f in the matrix entries is not obvious, the proposition nevertheless guarantees the existence of f . From this we can conclude, for example, that the set of $n \times n$ matrices A with a repeated eigenvalue (i.e., those for which p vanishes) forms a variety within $M_n(F)$. This exemplifies the potency of symmetric polynomials in the eigenvalues.

(iii) The reader may wish to discover an explicit expression for a polynomial map f on the entries of a 3×3 matrix that corresponds to the symmetric map p in (ii) above. □

Example 7.1.12

If $GL_n(F)$ is the *general linear group* of all invertible $n \times n$ matrices over F , is the inverse map $GL_n(F) \rightarrow GL_n(F)$, $A \mapsto A^{-1}$ a polynomial map of affine varieties in \mathbb{A}^{n^2} ? The answer is “no,” on two counts. First $GL_n(F)$ is not an affine variety according to our definition—it is determined by the *nonvanishing* of a polynomial, namely the determinant. (The reader may ask if it could still be determined by the zeros of some other polynomials. We will see later in Section 7.4 that this is not possible, because $GL_n(F)$ is a proper Zariski open subset of an irreducible variety.) Second, the inverse map is not polynomial in the entries of A because $A^{-1} = (1/\det A) \text{adj } A$, which is a “rational map” (although $\text{adj } A$ is polynomial in the entries of A). □

Remark 7.1.13

Rational maps (of the form $a \mapsto (p_1(a)/q_1(a), \dots, p_m(a)/q_m(a))$, where $p_i, q_i \in F[x_1, \dots, x_n]$ and $a \in \mathbb{A}^n$) play an important role in algebraic geometry (e.g., they are indispensable in projective varieties because polynomial maps don't make sense there), but they need to be handled with care. To keep our treatment as simple as possible, we will mostly avoid them. Likewise, although $GL_n(F)$ can be treated as a “quasi-affine variety” we shall sidestep this by formulating later arguments in terms of the affine variety $SL_n(F)$ whenever possible. Notice that the inverse map $A \mapsto A^{-1}$ is now a polynomial map of $SL_n(F)$ regarded as an affine variety in \mathbb{A}^{n^2} . \square

7.2 THE ZARISKI TOPOLOGY ON AFFINE n -SPACE

In this section we introduce the Zariski topology on affine n -space \mathbb{A}^n . When a family of sets are endowed with new topologies, one of the first questions one asks is which functions become continuous with respect to these topologies. In our case, the functions of interest are those mapping some \mathbb{A}^n into another \mathbb{A}^m . The Zariski topology turns out to be characterized as the weakest topology⁵ (that is, having the fewest open subsets) on the \mathbb{A}^n for which points are closed and polynomial maps are continuous. This topology is not always easy to visualize. (In the case of the real or complex field F , one normally relates the “pictures” to the standard metric topology.) Nevertheless, the Zariski topology is very useful, particularly with elementary topological arguments such as a polynomial function being completely determined by its values on a Zariski dense subset, and so on. These can often be used in “purely algebraic” situations. And, of course, the topology often suggests useful invariants under homeomorphism.

Rather than referring to an “affine variety V in \mathbb{A}^n ,” it is easier to say V is a **subvariety** of \mathbb{A}^n . The **Zariski topology** on affine n -space \mathbb{A}^n is defined by taking the subvarieties as its closed subsets. It is easy to check that the four axioms for the closed subsets of a topological space are satisfied:

- (1) The whole space is closed: $\mathbb{A}^n = V(0)$.
- (2) The empty set is closed: $\emptyset = V(1)$.
- (3) Finite unions of closed sets are closed:

$$V(S_1) \cup V(S_2) = V(S_1 S_2),$$

$$\text{where } S_1 S_2 = \{f_1 f_2 : f_1 \in S_1, f_2 \in S_2\}.$$

5. Serge Lang once commented that comparing two topologies using terms such as “weaker versus stronger” or “coarser versus finer” is ultimately confusing—no one is ever able to remember which is which!

(4) Arbitrary intersections of closed sets are closed:

$$\bigcap_{\gamma \in \Gamma} V(S_\gamma) = V\left(\bigcup_{\gamma \in \Gamma} S_\gamma\right).$$

Only the containment $V(S_1 S_2) \subseteq V(S_1) \cup V(S_2)$ in part (3) requires a little thought. Suppose $a \in V(S_1 S_2)$ but $a \notin V(S_1) \cup V(S_2)$. Then $f(a) \neq 0$ for some $f \in S_1$, and $g(a) \neq 0$ for some $g \in S_2$. But now $fg(a) = f(a)g(a) \neq 0$ contradicts $a \in V(S_1 S_2)$.

By definition, the Zariski open subsets are the complements of the Zariski closed subsets. They satisfy the four axioms for the open subsets of a topological space.⁶ Namely, \mathbb{A}^n is open, \emptyset is open, finite intersections of open sets are open, and arbitrary unions of open subsets are open. One useful observation is the following. For a polynomial $f \in F[x_1, x_2, \dots, x_n]$, let

$$U(f) = \{a \in \mathbb{A}^n : f(a) \neq 0\}$$

be the *nonvanishing set* of f . Then we have:

Proposition 7.2.1

The subsets $U(f)$, as f ranges over all polynomials from $F[x_1, x_2, \dots, x_n]$, form a basis of open subsets for the Zariski topology on \mathbb{A}^n .

Proof

The statement simply says that the $U(f)$ are open subsets and every open subset U is a union of some of these. Clearly, $U(f)$ is the complement of $V(f)$, whence open. Also, if U is the complement of some $V(S)$, then U is the union of the $U(f)$ for $f \in S$. \square

Example 7.2.2

Over a field F , a nonzero polynomial $f \in F[x]$ has only finitely many zeros. Hence, the closed subsets of \mathbb{A}^1 are arbitrary finite subsets and the whole space. The open subsets are the empty set and arbitrary cofinite subsets (i.e., complements of finite subsets). In particular, for $F = \mathbb{R}$ or \mathbb{C} , the Zariski topology is very much weaker than the standard metric topology (also known as the Euclidean topology). For \mathbb{A}^2 , it can be shown that the closed subsets are finite unions of isolated points and hypersurfaces determined by irreducible polynomials $f \in F[x_1, x_2]$. \square

6. A topology is traditionally presented by declaring its open subsets. In our case, it is simpler to declare the closed subsets.

Example 7.2.3

Let r and n be integers with $0 \leq r \leq n$ and $0 < n$. Then the set

$$V = \{A \in M_n(F) : \text{rank } A \leq r\}$$

is a Zariski closed subset of $M_n(F)$. (Here we are identifying $M_n(F)$ with \mathbb{A}^{n^2} .) In fact, since the rank of A is the largest integer s such that some $s \times s$ submatrix of A has nonzero determinant, we see that V is the subvariety of $M_n(F)$ determined by the following set S of polynomials in n^2 variables $x_{11}, x_{12}, \dots, x_{nn}$:

$$S = \{ \det B : B \text{ is an } (r+1) \times (r+1) \text{ submatrix of } (x_{ij}) \}.$$

(If $r = n$, then $S = \emptyset$.) For instance, for 3×3 matrices, the set of matrices with rank at most 1 is the variety determined by the following nine polynomials in the entries of the 9-tuple $\mathbf{x} = (x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}, x_{31}, x_{32}, x_{33})$:

$$\begin{aligned} p_1(\mathbf{x}) &= x_{11}x_{22} - x_{21}x_{12}, & p_2(\mathbf{x}) &= x_{11}x_{23} - x_{21}x_{13}, & p_3(\mathbf{x}) &= x_{12}x_{23} - x_{22}x_{13}, \\ p_4(\mathbf{x}) &= x_{11}x_{32} - x_{31}x_{12}, & p_5(\mathbf{x}) &= x_{11}x_{33} - x_{31}x_{13}, & p_6(\mathbf{x}) &= x_{12}x_{33} - x_{32}x_{13}, \\ p_7(\mathbf{x}) &= x_{21}x_{32} - x_{31}x_{22}, & p_8(\mathbf{x}) &= x_{21}x_{33} - x_{31}x_{23}, & p_9(\mathbf{x}) &= x_{22}x_{33} - x_{32}x_{23}. \end{aligned}$$

Again, with this type of example, we don't usually need to know exactly what the polynomials are, only that they exist.

Of course, the complementary set

$$U = \{A \in M_n(F) : \text{rank } A > r\}$$

is therefore a Zariski open subset. With this pair of examples of a Zariski closed and a Zariski open subset, the inequalities \leq and $>$ match our intuition of what should be closed and what should be open, based on our experience with the Euclidean topologies. But one must always be careful here, otherwise one might think the sets

$$W = \{A \in M_n(F) : \text{rank } A \geq r\},$$

$$X = \{A \in M_n(F) : \text{rank } A < r\}$$

should be respectively Zariski closed and Zariski open. But that is completely round the wrong way! For W corresponds to U when we replace r by $r-1$, so W is Zariski open. And X corresponds to V when we replace r by $r-1$, so X is Zariski closed.

The reader may risk further confusion by showing that the set

$$Y = \{A \in M_2(F) : \text{rank } A = 1\}$$

is neither Zariski closed nor Zariski open in $M_2(F)$ when F is infinite! (The reason for this will become clearer in Section 7.4.) \square

Example 7.2.4

Is the set

$$Z = \{A \in M_n(F) : A \text{ is 1-regular}\}$$

a Zariski open or a Zariski closed subset of $M_n(F)$, or neither? Since a matrix $A \in M_n(F)$ is 1-regular if and only if

$$(*) \quad \dim E_\lambda = 1$$

for all the eigenspaces E_λ , one might guess that Z is Zariski closed. To clinch that argument one would need to show that $(*)$ is determined by some polynomial equations in the entries of A , and that is not clear. A better approach is to observe that 1-regularity is, by Proposition 1.1.2, equivalent to

$$(**) \quad I, A, A^2, \dots, A^{n-1} \text{ are linearly independent.}$$

If we form the $n^2 \times n$ matrix

$$M = \begin{bmatrix} | & | & | & & | \\ I & A & A^2 & \cdots & A^{n-1} \\ | & | & | & & | \end{bmatrix},$$

whose columns are the powers of A written as $n^2 \times 1$ column vectors, then $(**)$ is equivalent to M having rank n . By the same argument as in the previous example, this is a Zariski open condition on (the entries of) A because it says some $n \times n$ submatrix of M has nonzero determinant. Hence, our first guess was wrong, and Z is in fact a Zariski open subset. \square

Notice that a singleton set $\{(a_1, a_2, \dots, a_n)\}$ is Zariski closed in \mathbb{A}^n because it is the subvariety $V(\{x_1 - a_1, x_2 - a_2, \dots, x_n - a_n\})$. Thus, the Zariski topology satisfies the so-called T_1 separation axiom.⁷ However, the Hausdorff separation axiom T_2 (the ability to “house off” two distinct points in disjoint neighbourhoods) fails badly because of our next proposition. But first a lemma.

We know that a nonzero polynomial $f(x) \in F[x]$ in one variable and, over a field F , can vanish at only finitely many elements of F . (In fact, the number of roots can’t exceed the degree of $f(x)$.) But of course a nonzero polynomial $f(x_1, \dots, x_n) \in F[x_1, \dots, x_n]$ in more than one variable, and over

7. In consequence, over a finite field F , the Zariski topology on \mathbb{A}^n is just the discrete topology. For this reason, we often assume F is infinite.

an infinite field F , can vanish at an infinite number of points in \mathbb{A}^n . Nevertheless we have:

Lemma 7.2.5

Let F be an infinite field and $f(x_1, \dots, x_n)$ a nonzero polynomial from $F[x_1, \dots, x_n]$. Then there are an infinite number of points in \mathbb{A}^n at which f does not vanish.⁸

Proof

We proceed by induction on n , the case $n = 1$ being clear. Suppose $n > 1$. We can assume f is not a constant, and without loss of generality, f has positive degree in x_n . Write

$$f(x_1, \dots, x_n) = f_0 + f_1 x_n + f_2 x_n^2 + \dots + f_k x_n^k$$

where the $f_i \in F[x_1, \dots, x_{n-1}]$, $k \geq 1$, and f_k is nonzero. By induction, we can suppose $f_k(a_1, \dots, a_{n-1}) \neq 0$ for some $(a_1, \dots, a_{n-1}) \in \mathbb{A}^{n-1}$. Now the single variable polynomial

$$\begin{aligned} g(x) &= f(a_1, \dots, a_{n-1}, x) \in F[x] \\ &= f_0(a_1, \dots, a_{n-1}) + f_1(a_1, \dots, a_{n-1})x + \dots + f_k(a_1, \dots, a_{n-1})x^k \end{aligned}$$

is nonzero and so has only finitely many zeros, say z_1, \dots, z_m . Inasmuch as F is infinite, this allows an infinite choice of points $(a_1, \dots, a_{n-1}, b) \in \mathbb{A}^n$ at which f does not vanish, namely those for which $b \neq z_1, \dots, z_m$. \square

Proposition 7.2.6

Suppose F is an infinite field. Then any two nonempty Zariski open subsets of \mathbb{A}^n have a nonempty intersection.

Proof

By Proposition 7.2.1, it suffices to show $U(f)$ and $U(g)$ intersect when they are nonempty. Now this requires f and g to be nonzero, whence fg is nonzero because $F[x_1, x_2, \dots, x_n]$ is an integral domain. Since F is infinite, by Lemma 7.2.5 we know fg can't vanish everywhere. Hence, $\emptyset \neq U(fg) \subseteq U(f) \cap U(g)$. \square

Recall that a mapping $f : X \rightarrow Y$ from a topological space X into a topological space Y is **continuous** (relative to the given topologies) if the

8. Polynomials in $F[x_1, \dots, x_n]$ over an infinite field F are therefore faithfully represented by their corresponding polynomial functions. This fails for a finite field F of order q (necessarily a prime power), because the nonzero polynomial $x^q - x \in F[x]$ induces the zero polynomial map $F \rightarrow F$, $a \mapsto a^q - a$.

inverse image $f^{-1}(U) = \{x \in X : f(x) \in U\}$ of every open subset U of Y is an open subset of X . The mapping f is a **homeomorphism** if f is a bijection and both f and f^{-1} are continuous (so f induces a bijection between the families of open subsets of X and Y , respectively).

Proposition 7.2.7

Every polynomial map $f : \mathbb{A}^n \rightarrow \mathbb{A}^m$ is continuous with respect to the Zariski topologies.

Proof

Suppose there are polynomials $p_1, p_2, \dots, p_m \in F[x_1, x_2, \dots, x_n]$ such that for all $a \in \mathbb{A}^n$, $f(a) = (p_1(a), p_2(a), \dots, p_m(a))$. By Proposition 7.2.1, to show f is continuous, it is enough to show that $f^{-1}(U(g))$ is open for all $g \in F[x_1, x_2, \dots, x_m]$. In turn, we need only show that if $a \in f^{-1}(U(g))$, then $a \in U(h) \subseteq f^{-1}(U(g))$ for some $h \in F[x_1, x_2, \dots, x_n]$. This just says that $h(a) \neq 0$, and that $h(b) \neq 0$ implies $g(f(b)) \neq 0$. So we can take $h(x_1, \dots, x_n) = g(p_1(x_1, \dots, x_n), \dots, p_m(x_1, \dots, x_n))$. \square

Remark 7.2.8

Suppose someone has endowed the affine spaces \mathbb{A}^n with new topologies such that points are closed and polynomial maps are continuous. Let $f \in F[x_1, x_2, \dots, x_n]$ and consider the associated polynomial map $f : \mathbb{A}^n \rightarrow \mathbb{A}^1$. Since $\{0\}$ is closed, and f is continuous, the inverse image of the complement of $\{0\}$ must be open in \mathbb{A}^n . That is, $U(f)$ is open. By Proposition 7.2.1, all Zariski open subsets are open in the new topology. This justifies our earlier opening statement that the Zariski topology is the weakest for which points are closed and polynomial maps are continuous. In particular, since points are closed and polynomial maps are continuous in the standard topology when $F = \mathbb{R}$ or \mathbb{C} , the Zariski topology is (very much) weaker than the standard topology. One should also observe that over a field F , the Zariski topology on \mathbb{A}^n for $n > 1$ is not the product topology on n copies of \mathbb{A}^1 with its Zariski topology. For example, the Zariski open subset of \mathbb{R}^2 determined by $x + y \neq 0$ is not the union of open subsets of the form $O_1 \times O_2$ where both O_i are Zariski open in \mathbb{R} (because such nonempty O_i must be cofinite). \square

Example 7.2.9

Any function $f : \mathbb{A}^1 \rightarrow \mathbb{A}^1$ for which the “fibres” $f^{-1}(y)$ are finite, for all $y \in \mathbb{A}^1$, is Zariski continuous. We just need to check that inverse images of closed sets are closed. But the closed sets are the finite subsets and the whole space, so that is trivial. In particular, there are Zariski continuous functions that are not polynomial maps (cf. Remark 7.2.8). For instance, the map $f : \mathbb{R} \rightarrow \mathbb{R}$ given by

$$f(x) = \begin{cases} x^2 + 1 & \text{if } x \text{ is rational} \\ x^2 & \text{if } x \text{ is irrational} \end{cases}$$

is continuous in the Zariski topology but nowhere continuous in the standard topology.

By the same argument, the mapping in Example 7.1.4(ii) is a Zariski homeomorphism (if the second variety has the induced Zariski topology). However, we will see later in Example 7.7.5 that the two affine varieties V and W in Example 7.1.4(ii) have different algebraic-geometric properties. \square

7.3 THE THREE THEOREMS UNDERPINNING BASIC ALGEBRAIC GEOMETRY

There are three theorems that every student of even elementary algebraic geometry should be aware of. They provide the foundations for most of the theory. All three have simple statements in terms of commutative rings, and their proofs, although not easy, can be understood by a graduate student who knows the basics of commutative rings and fields. Not surprisingly, the three theorems carry impressive monikers: Hilbert's basis theorem, Hilbert's nullstellensatz, and Noether's normalization theorem. We refer the interested reader to *Basic Algebra II* by Jacobson, *Algebraic Geometry* by Milne, or *Undergraduate Commutative Algebra* by Reid, for the full details of their proofs.

Theorem 7.3.1 (Hilbert's Basis Theorem)

Let F be any field. Then every ideal I of the polynomial ring $R = F[x_1, x_2, \dots, x_n]$ is finitely generated. That is, there exist $f_1, f_2, \dots, f_k \in I$ such that $I = f_1R + f_2R + \dots + f_kR$. Equivalently, every ascending chain of ideals of R

$$I_1 \subseteq I_2 \subseteq \dots \subseteq I_j \subseteq I_{j+1} \subseteq \dots$$

eventually becomes stationary: for some j , we must have $I_j = I_{j+1} = I_{j+2} = \dots$.

Remark 7.3.2

The stationary property of ascending chains of ideals of R is called the "ascending chain condition" (acc). A ring that satisfies the acc is called "Noetherian." A corollary to the basis theorem is that all finitely generated commutative algebras over F are Noetherian, because they are homomorphic images of some $F[x_1, x_2, \dots, x_n]$. \square

Theorem 7.3.3 (Hilbert's Nullstellensatz)

Let F be an algebraically closed field. Let I be a proper ideal of the polynomial ring $F[x_1, x_2, \dots, x_n]$. Then there is a common zero $a = (a_1, a_2, \dots, a_n) \in \mathbb{A}^n$ of all the polynomials in I , i.e., $f(a) = 0$ for all $f \in I$.

Remarks 7.3.4

- (1) “Nullstellensatz” befittingly means “zero-point theorem” in German.
- (2) When $n = 1$, the theorem is equivalent to F being algebraically closed because nonzero proper ideals of $R = F[x]$ take the form fR where f is a nonconstant polynomial. In particular, over the complex field \mathbb{C} , Hilbert’s nullstellensatz can be viewed as a generalization of the fundamental theorem of algebra.
- (3) The requirement that I not be all of $F[x_1, x_2, \dots, x_n]$ is clearly necessary. For if $f_1g_1 + f_2g_2 + \dots + f_kg_k$ is the constant polynomial 1 for some polynomials g_1, g_2, \dots, g_k , then f_1, f_2, \dots, f_k can’t have a common zero.
- (4) An equivalent statement of the theorem is that the maximal ideals of $R = F[x_1, x_2, \dots, x_n]$ take the form

$$(x_1 - a_1)R + (x_2 - a_2)R + \dots + (x_n - a_n)R,$$

where $(a_1, a_2, \dots, a_n) \in \mathbb{A}^n$. This enables one to establish a natural homeomorphism of \mathbb{A}^n with its Zariski topology and the space of maximal ideals of $F[x_1, x_2, \dots, x_n]$ with its topology. We won’t develop this connection but the interested reader can pursue the details in Hulek’s *Elementary Algebraic Geometry*, Chapter 1.

- (5) We shall mention another important equivalent statement of the Nullstellensatz in Theorem 7.7.2. □

Theorem 7.3.5 (Noether’s Normalization Theorem: Version 1)

Let F be an infinite field and let A be a finitely generated commutative algebra over F . Then there exist elements $y_1, y_2, \dots, y_m \in A$ such that:

- (1) y_1, y_2, \dots, y_m are algebraically independent over F , that is, the only polynomial $f \in F[x_1, x_2, \dots, x_m]$ for which $f(y_1, y_2, \dots, y_m) = 0$ is the zero polynomial;
- (2) if B is the subalgebra of A generated by y_1, y_2, \dots, y_m , then A is finitely generated as a module over B , that is, $A = a_1B + a_2B + \dots + a_nB$ for some $a_i \in A$.

Theorem 7.3.6 (Noether’s Normalization Theorem: Version 2)

Let F be an algebraically closed field and let V be an affine variety in \mathbb{A}^n with the property that the ideal $I = \{f \in F[x_1, x_2, \dots, x_n] : f(a) = 0 \text{ for all } a \in V\}$ is a prime ideal of $F[x_1, x_2, \dots, x_n]$. (This means V is an “irreducible” variety in the sense of the next section.) Then for some $m \leq n$, there exists a polynomial map

$$\phi : V \rightarrow \mathbb{A}^m$$

from V **onto** \mathbb{A}^m such that the fibres $\phi^{-1}(b)$ are finite for all $b \in \mathbb{A}^m$. (We say such a map ϕ is **quasi-finite**.⁹)

Remarks 7.3.7

- (1) Some authors refer to version 2 as “the” Noether normalization theorem, while others give that honor to version 1, and they then regard version 2 as the geometric interpretation of it. We have decided to cover both bases.
- (2) In the case of a vector subspace V of F^n , we can parameterize the elements of V by m -tuples (for $m = \dim V$). Such a parametrization is not possible for irreducible subvarieties V of \mathbb{A}^n . The second version of Noether’s theorem is the best we can do. It turns out that the integer m in the theorem is the algebraic geometry “dimension” of the variety V , discussed in Section 7.8. □

7.4 IRREDUCIBLE VARIETIES

An affine variety V in affine n -space \mathbb{A}^n is called *irreducible* if V is not a union of two proper subvarieties: $V = V_1 \cup V_2$ for subvarieties V_1, V_2 of \mathbb{A}^n implies $V_1 = V$ or $V_2 = V$. The notion of irreducibility makes sense in any topological space. A topological space Y is **irreducible** if Y is not the union of two proper closed subsets, equivalently, any two nonempty open subsets of Y have nonempty intersection. Moreover, we can talk of an arbitrary *subset* X of Y being irreducible relative to the induced topology on X . (Recall that in this topology, the open subsets of X are the $U \cap X$ where U is any open subset of Y .) Although our primary interest is in irreducible varieties, for our purposes it makes life easier to treat irreducibility in this more general subset context.

Definition 7.4.1: A subset X of \mathbb{A}^n is **irreducible** if X is irreducible as a topological space relative to the induced Zariski topology: $X \neq X_1 \cup X_2$ for any proper closed subsets X_1, X_2 of X . Of course, we say X is **reducible** if it is not irreducible.

Knowing that a variety of matrices is irreducible can often be very useful in linear algebra problems, even when the statement of the problem seems to have no connection with algebraic geometry. That is really what this chapter is about. But we must point out that, in general, it can be difficult to determine whether a given variety is irreducible.

Our first example of an irreducible variety is simple but crucial and will be called on frequently.

9. The term “finite map” is reserved for something slightly stronger, namely, a polynomial mapping $\phi : V \rightarrow W$ such that the co-ordinate ring $F[V]$ is a finitely generated $F[W]$ -module via the induced F -algebra homomorphism $f^* : F[W] \rightarrow F[V]$. See Section 7.7 for these notions.

Proposition 7.4.2

Over an infinite field F , \mathbb{A}^n is irreducible.

Proof

This is immediate from Proposition 7.2.6. □

Recall that for a subset X of a topological space Y , there is a unique smallest closed subset of Y containing X (namely, the intersection of all closed subsets containing X). This is called the **closure** of X , and we denote it by \overline{X} . The elements b of \overline{X} are characterized by the property that for every open subset U of Y containing b , we must have $U \cap X \neq \emptyset$. A subset X whose closure is all of Y is said to be a **dense** subset of Y (relative to the given topology). The following result explains why in arguments involving irreducible subsets X of \mathbb{A}^n , one can nearly always assume that X is in fact an irreducible variety.

Proposition 7.4.3

A subset X of \mathbb{A}^n is irreducible if and only if its Zariski closure \overline{X} is an irreducible subvariety of \mathbb{A}^n .

Proof

Assume X is irreducible and $\overline{X} = W_1 \cup W_2$ where each W_i is closed in \overline{X} . Then $X = (X \cap W_1) \cup (X \cap W_2)$ and so, since each $X \cap W_i$ is closed in X , we have $X = X \cap W_1$ or $X = X \cap W_2$, say the former. Then $X \subseteq W_1$, whence $\overline{X} = W_1$. This shows \overline{X} is irreducible. The converse is similar, using the fact that the closure of the union of two subsets is the union of their closures. □

Corollary 7.4.4

If V is an irreducible subset of \mathbb{A}^n , then a nonempty Zariski open subset U of V is also irreducible.

Proof

Since V is irreducible, U has a nonempty intersection with each nonempty open subset of V , whence U is dense in V . Thus, $\overline{U} = \overline{V}$, whence by Proposition 7.4.3, \overline{U} is an irreducible subvariety of \mathbb{A}^n . Therefore, again by the proposition, U is an irreducible subset of \mathbb{A}^n . □

Remark 7.4.5

Sometimes this corollary is the quickest way of verifying certain subsets of \mathbb{A}^n are irreducible. For instance, in the matrix setting where we identify $M_n(F)$ with \mathbb{A}^{n^2} , since $GL_n(F)$ is the nonvanishing set of the polynomial map $\det: \mathbb{A}^{n^2} \rightarrow \mathbb{A}^1$, we see that $GL_n(F)$ is a nonempty Zariski open subset of the variety $M_n(F)$, whence by Proposition 7.4.2 and Corollary 7.4.4 we have $GL_n(F)$ is an irreducible subset when F is infinite. □

Corollary 7.4.6

Over an infinite field F , every cofinite subset X of \mathbb{A}^n is irreducible.

Proof

Inasmuch as singleton sets are closed, the cofinite subsets are nonempty open subsets of the irreducible \mathbb{A}^n and so irreducible by Corollary 7.4.4. \square

The following basic property is most useful and will be used frequently (sometimes without explicit mention).¹⁰

Proposition 7.4.7

Let V be an affine subvariety of \mathbb{A}^n and suppose that $f : V \rightarrow \mathbb{A}^m$ and $g : V \rightarrow \mathbb{A}^m$ are polynomial functions from V into some affine space \mathbb{A}^m . If f and g agree on some Zariski dense subset Z of V , then $f = g$.

Proof

The subset $Y = \{v \in V : f(v) = g(v)\}$ is closed in V because it is the inverse image of the closed subset $\{0\}$ under the Zariski continuous function $f - g$. Since Y contains the dense subset Z , we must have $Y = V$. Thus, f and g agree everywhere. \square

The ring $F[x_1, x_2, \dots, x_n]$ has no zero divisors. The following result says that irreducibility of a subvariety V of \mathbb{A}^n is characterized by the condition that the ring of polynomial functions on V has no zero divisors. In Section 7.7, we shall give a smarter statement of this result in terms of the co-ordinate ring of the variety. For simple varieties, the condition provides a quick way of verifying irreducibility.

Proposition 7.4.8

Let V be a subvariety of \mathbb{A}^n . Then V is irreducible if and only if, for all $f, g \in F[x_1, x_2, \dots, x_n]$,

$$fg = 0 \text{ on } V \implies f = 0 \text{ on } V \text{ or } g = 0 \text{ on } V.$$

Proof

By Proposition 7.2.1, the sets $X_f = U(f) \cap V = \{v \in V : f(v) \neq 0\}$, as f ranges over $F[x_1, x_2, \dots, x_n]$, form a basis of open subsets of V (in its induced topology). Hence, V is irreducible precisely when nonempty X_f, X_g give $X_{fg} = X_f \cap X_g$ nonempty, that is, if and only if whenever $fg = 0$ on V then either f or g is zero on V . \square

10. It is not true that general Zariski continuous functions $f : V \rightarrow W, g : V \rightarrow W$ that agree on some Zariski dense subset of V must agree everywhere! For instance, let $V = W = \mathbb{A}^1$ and consider distinct bijections f, g that agree on some cofinite subset of V .

To make further progress with irreducible sets, we need the Hilbert basis theorem from Section 7.3. As a moment's thought will confirm, the variety $V = V(S)$ determined by a subset S of $F[x_1, x_2, \dots, x_n]$ is the same as the variety $V(I)$ determined by the ideal I of $F[x_1, x_2, \dots, x_n]$ generated by S . By the Hilbert Basis Theorem 7.3.1, I is finitely generated, say by polynomials f_1, f_2, \dots, f_k . Now $V = V(\{f_1, f_2, \dots, f_k\})$. Therefore, every variety is determined by some finite set of polynomials.

We have a correspondence $J \mapsto V(J)$ between the set of ideals J of $F[x_1, \dots, x_n]$ and the set of subvarieties W of \mathbb{A}^n . There is a correspondence also in the other direction: given a subvariety W of \mathbb{A}^n , let

$$I(W) = \{f \in F[x_1, x_2, \dots, x_n] : f(w) = 0 \text{ for all } w \in W\},$$

which is clearly an ideal. We record the following key features of these correspondences:

Proposition 7.4.9

The maps I and V satisfy the following for any subvarieties X, Y of \mathbb{A}^n and ideals J, K of $F[x_1, x_2, \dots, x_n]$:

- (1) If $J \subseteq K$, then $V(K) \subseteq V(J)$.
- (2) If $X \subseteq Y$, then $I(X) \supseteq I(Y)$.
- (3) $J \subseteq I(V(J))$.
- (4) $X = V(I(X))$.

Proof

The only item that needs checking is (4), and then only the containment $V(I(X)) \subseteq X$. Since X is a subvariety, $X = V(J)$ for some ideal J . Using (3), we get $J \subseteq I(V(J)) = I(X)$ and so by (1) we have $V(I(X)) \subseteq V(J) = X$. \square

Proposition 7.4.10

In affine n -space \mathbb{A}^n , every descending chain of closed subsets becomes stationary (equivalently, every ascending chain of open subsets becomes stationary). The same is true for descending chains of closed subsets of any subset X of \mathbb{A}^n relative to the induced Zariski topology.

Proof

Suppose $X_1 \supseteq X_2 \supseteq \dots \supseteq X_i \supseteq X_{i+1} \supseteq \dots$ is a descending chain of closed subsets (that is, subvarieties) X_i . Applying the Hilbert Basis Theorem 7.3.1 to the ascending chain of ideals

$$I(X_1) \subseteq I(X_2) \subseteq \dots \subseteq I(X_i) \subseteq I(X_{i+1}) \subseteq \dots$$

obtained from Proposition 7.4.9 (2), we get $I(X_j) = I(X_{j+1}) = \cdots$ for some j . By Proposition 7.4.9 (4), $X_j = V(I(X_j)) = V(I(X_{j+1})) = X_{j+1} = \cdots$. This establishes the claim for \mathbb{A}^n .

If $X_1 \supseteq X_2 \supseteq \cdots \supseteq X_i \supseteq X_{i+1} \supseteq \cdots$ is a descending chain of closed subsets of X , then we have $\overline{X}_1 \supseteq \overline{X}_2 \supseteq \cdots \supseteq \overline{X}_i \supseteq \overline{X}_{i+1} \supseteq \cdots$, whence $\overline{X}_j = \overline{X}_{j+1} = \cdots$ for some j by the established property for \mathbb{A}^n . Now $X_j = \overline{X}_j \cap X = \overline{X}_{j+1} \cap X = X_{j+1} = X_{j+2} = \cdots$. \square

Proposition 7.4.11

Every subset X of \mathbb{A}^n has a decomposition

$$X = X_1 \cup X_2 \cup \cdots \cup X_k$$

as a union of irreducible closed subsets X_i of X with $X_i \not\subseteq X_j$ for $i \neq j$. This decomposition is unique up to the order of the terms.

Proof

The proof works in any topological space that has the descending chain condition on closed subsets. First we establish the existence of such a decomposition. By Proposition 7.4.10, every nonempty collection \mathcal{C} of closed subsets of \mathbb{A}^n must have a minimal member, that is, one not properly containing any other member of \mathcal{C} (otherwise we could construct an infinite strictly descending chain of closed subsets). Let \mathcal{C} be the collection of all closed subsets Y of \mathbb{A}^n that fail to have a decomposition into irreducible closed subsets. Suppose \mathcal{C} is nonempty. Let Y be a minimal member of \mathcal{C} . Then Y can't be irreducible, so there is a decomposition $Y = Y_1 \cup Y_2$ where Y_1, Y_2 are proper closed subsets of Y . By minimality of Y , we must have $Y_1, Y_2 \notin \mathcal{C}$. Therefore Y_1 and Y_2 have decompositions into irreducible closed subsets. Combining them gives a decomposition for Y as a union of irreducible closed subsets, a contradiction. Hence, \mathcal{C} is empty, that is, every closed subset of \mathbb{A}^n has the desired decomposition. Now given an arbitrary subset X of \mathbb{A}^n , its Zariski closure Y has a decomposition into irreducible closed subsets Y_1, Y_2, \dots, Y_k . Taking $X_i = X \cap Y_i$ then gives X as a finite union of irreducible closed subsets X_i of X . Discarding any redundant terms, that is any X_i contained in some other X_j , produces the desired decomposition.

To establish uniqueness, suppose

$$X = X_1 \cup X_2 \cup \cdots \cup X_k = Y_1 \cup Y_2 \cup \cdots \cup Y_m$$

are two decompositions into irreducible closed subsets with $X_i \not\subseteq X_j$ and $Y_i \not\subseteq Y_j$ for $i \neq j$. Fix i with $1 \leq i \leq k$. Then

$$X_i = X_i \cap X = \bigcup_{j=1}^m (X_i \cap Y_j)$$

is a decomposition of X_i into closed subsets, so by the irreducibility of X_i we have $X_i = X_i \cap Y_j$ for some j , that is $X_i \subseteq Y_j$. By the same argument, but changing the roles of the two decompositions, we have $Y_j \subseteq X_l$ for some l , which implies $X_i \subseteq X_l$ and hence $i = l$. Thus, $X_i = Y_j$. Hence, each member of the first decomposition occurs (exactly once) in the second, and vice versa. So apart from the order of the terms, the two decompositions are the same. \square

We refer to the unique X_1, X_2, \dots, X_k in Proposition 7.4.11 as the **irreducible components** of X .

Remark. If we didn't insist that the X_i be closed subsets of X , only irreducible, we wouldn't have uniqueness. For instance, if F is an infinite field then cofinite subsets of \mathbb{A}^n are irreducible by Corollary 7.4.6 and so, for distinct $a, b \in \mathbb{A}^n$, we have the following four different decompositions of \mathbb{A}^n into irreducible subsets:

$$\begin{aligned}\mathbb{A}^n &= \mathbb{A}^n \\ &= \{a\} \cup (\mathbb{A}^n \setminus \{a\}) \\ &= \{b\} \cup (\mathbb{A}^n \setminus \{b\}) \\ &= \{a\} \cup \{b\} \cup (\mathbb{A}^n \setminus \{a, b\}).\end{aligned}$$

Only the first is the true decomposition into irreducible components. \square

Example 7.4.12

We show in Corollary 7.7.3 that, over an algebraically closed field F , if f is a nonconstant polynomial in $F[x_1, x_2, \dots, x_n]$, and f_1, f_2, \dots, f_k are its distinct irreducible factors, then the irreducible decomposition of the hypersurface $V(f)$ determined by f is

$$V(f) = V(f_1) \cup V(f_2) \cup \dots \cup V(f_k).$$

Sometimes this decomposition will also give the irreducible components when the field F is not algebraically closed, but caution should be exercised. For instance, the irreducible components of the real affine variety in \mathbb{A}^2 determined by the polynomial $f(x, y) = (y - x - 2)^2(y - x^2 - 1)$ are indeed the line $y = x + 2$ and the parabola $y = x^2 + 1$, as in Figure 7.4 on the next page. Certainly the line is irreducible and so too is the parabola because the mapping $t \mapsto (t, t^2 + 1)$ is an affine isomorphism of \mathbb{A}^1 onto the parabola. (In fact, we can deduce irreducibility of a variety from any polynomial parameterization in view of our next proposition.) However, for the real polynomial $f(x, y) = x^2(x - 1)^2 + y^2$, it is already factored $f = f_1$ into real irreducibles but $V(f_1) = \{(0, 0)\} \cup \{(1, 0)\}$ is not irreducible. \square

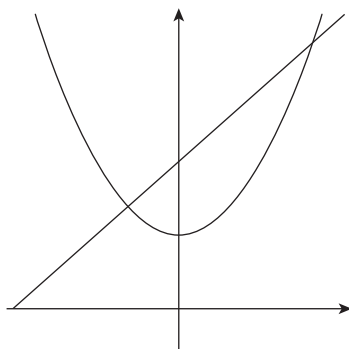


Figure 7.4 $f(x, y) = (y - x - 2)^2 (y - x^2 - 1)$

Example 7.4.13

When F is infinite, what are the irreducible components of the variety

$$V = \{A \in M_2(F) : A^2 = A\}$$

of 2×2 idempotent matrices? Note that the only nonsingular idempotent matrix is the identity I . Thus, V decomposes into proper subvarieties

$$V = V_0 \cup V_1,$$

where V_0 is the set of singular idempotents and $V_1 = \{I\}$. Hence, V is reducible. Clearly V_1 is irreducible, but $V_0 = \{0\} \cup W$ where W is the set of rank 1 idempotents. Since the characteristic polynomial of a 2×2 matrix A is $p(x) = x^2 - (\text{tr } A)x + \det A$, we see that

$$W = \{A \in M_2(F) : \det A = 0, \text{tr } A = 1\}.$$

Thus, W is the subvariety of $M_2(F)$ determined by the polynomials

$$f_1(x_{11}, x_{12}, x_{21}, x_{22}) = x_{11}x_{22} - x_{21}x_{12},$$

$$f_2(x_{11}, x_{12}, x_{21}, x_{22}) = x_{11} + x_{22} - 1.$$

Moreover, W is irreducible. One way to see this is to observe that

$$W = \bigcup_{T \in GL_2(F)} T^{-1} \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\} T$$

and to use the upcoming Proposition 7.4.18. Thus, the irreducible components of V are $\{0\}$, $\{I\}$, and W . \square

The next result also frequently presents a practical way of verifying that a variety is irreducible.

Proposition 7.4.14

Suppose X and Y are subsets of affine spaces \mathbb{A}^n and \mathbb{A}^m , respectively, and $\theta : X \rightarrow Y$ is a Zariski continuous mapping. If X is irreducible, then so is $\theta(X)$.

Proof

Let U and W be nonempty open subsets of $\theta(X)$ (in the induced Zariski topology). By continuity, $\theta^{-1}(U)$ and $\theta^{-1}(W)$ are nonempty open subsets of X , whence their intersection is nonempty because X is irreducible. Now $\emptyset \neq \theta(\theta^{-1}(U) \cap \theta^{-1}(W)) \subseteq U \cap W$, which implies $U \cap W$ is nonempty. This shows $\theta(X)$ is irreducible. \square

Corollary 7.4.15

Let X be an irreducible subset of $M_n(F)$ and let $T \in M_n(F)$. Then $TX = \{TA : A \in X\}$ is an irreducible subset of $M_n(F)$. Similarly, so is XT .

Proof

The map $\theta : X \rightarrow M_n(F)$, $A \mapsto TA$ is a polynomial map on the entries of A , hence Zariski continuous by Proposition 7.2.7. (Here we are identifying $M_n(F)$ with \mathbb{A}^{n^2} in the standard way.) Therefore, $TX = \theta(X)$ is irreducible. \square

If V and W are subsets of affine spaces \mathbb{A}^m and \mathbb{A}^n , respectively, we view their **product** $V \times W = \{(v, w) : v \in V, w \in W\}$ as a subset of \mathbb{A}^{m+n} in the obvious way. Note that if V and W are affine subvarieties, then so is $V \times W$. For suppose that $f_1, f_2, \dots, f_k \in F[x_1, x_2, \dots, x_m]$ determine V , and that $g_1, g_2, \dots, g_l \in F[x_1, x_2, \dots, x_n]$ determine W . For $j = 1, 2, \dots, l$, let $\bar{g}_j \in F[x_{m+1}, x_{m+2}, \dots, x_{m+n}]$ be the polynomial obtained from $g_j(x_1, x_2, \dots, x_n)$ by replacing x_i by x_{m+i} . Then $f_1, f_2, \dots, f_k, \bar{g}_1, \bar{g}_2, \dots, \bar{g}_l \in F[x_1, x_2, \dots, x_{m+n}]$ determine $V \times W$. Normally one wouldn't expect a "product" of irreducible algebraic objects to be irreducible, but that is what happens in our setting:

Proposition 7.4.16

If V and W are irreducible subsets of \mathbb{A}^m and \mathbb{A}^n , then $V \times W$ is an irreducible subset of \mathbb{A}^{m+n} .

Proof

Suppose $V \times W = Z_1 \cup Z_2$ is a decomposition into closed subsets. Fix $w \in W$. Then $V \times \{w\}$ is an irreducible subset of \mathbb{A}^{m+n} by Proposition 7.4.14 because the mapping $v \mapsto (v, w)$ is an isomorphism of V onto $V \times \{w\}$. Since

$$V \times \{w\} = ((V \times \{w\}) \cap Z_1) \cup ((V \times \{w\}) \cap Z_2)$$

decomposes the irreducible $V \times \{w\}$ into closed subsets, we must have

$$\begin{aligned} \text{either } (1) \quad V \times \{w\} &\subseteq Z_1 \\ \text{or } (2) \quad V \times \{w\} &\subseteq Z_2. \end{aligned}$$

Now let w range over W and let W_1 and W_2 be the subsets of W determined respectively by conditions (1) and (2). Then $W = W_1 \cup W_2$ so we are finished by the irreducibility of W if we can show the W_i are closed subsets of W , because then either $W_1 = W$ or $W_2 = W$, which in turn implies either $Z_1 = V \times W$ or $Z_2 = V \times W$. For each $v \in V$, the map $\theta_v : W \rightarrow V \times W$, $w \mapsto (v, w)$ is a polynomial map, hence Zariski continuous. Therefore, $\theta_v^{-1}(Z_1)$ is a closed subset of W because Z_1 is closed. But W_1 is the intersection of these closed sets as v ranges over V , so W_1 is closed. Similarly, so is W_2 . \square

As we earlier remarked, to keep our discussions simple we will generally avoid using rational maps. However, the following is straightforward and quite useful.

Proposition 7.4.17

Let f_1, f_2, \dots, f_m and g_1, g_2, \dots, g_m be polynomials in $F[x_1, x_2, \dots, x_n]$, and let D be a nonempty subset of \mathbb{A}^n on which none of the g_i has a zero. Then the rational map

$$\theta = \left(\frac{f_1}{g_1}, \frac{f_2}{g_2}, \dots, \frac{f_m}{g_m} \right) : D \longrightarrow \mathbb{A}^m$$

is Zariski continuous on D .

Proof

By Proposition 7.2.1, it is enough to show $\theta^{-1}(U(h))$ is open for the nonvanishing set $U(h)$ of a polynomial $h \in F[x_1, x_2, \dots, x_m]$. We can assume $\theta^{-1}(U(h))$ is nonempty. Let $d \in \theta^{-1}(U(h))$, in other words, $d \in D$ with $h(\theta(d)) \neq 0$. It is enough to find $f \in F[x_1, x_2, \dots, x_n]$ with $d \in U(f) \cap D \subseteq \theta^{-1}(U(h))$. That is, we want:

- (i) $f(d) \neq 0$,
- (ii) $f(b) \neq 0 \Rightarrow h(\theta(b)) \neq 0$ for all $b \in D$.

Let k be the degree of h and let

$$f(x_1, x_2, \dots, x_n) = (g_1(x_1, x_2, \dots, x_n))^k \cdots (g_m(x_1, x_2, \dots, x_n))^k h\left(\frac{f_1}{g_1}, \frac{f_2}{g_2}, \dots, \frac{f_m}{g_m}\right).$$

Note that f is a *polynomial* function in $F[x_1, x_2, \dots, x_n]$, not just a rational function. For $b = (b_1, b_2, \dots, b_n) \in D$, clearly $f(b) \neq 0$ exactly when $h(\theta(b)) \neq 0$. Thus, (i) and (ii) hold. \square

We will make good use of the following result. Its proof uses a rational map (and the previous proposition) but the result allows us to avoid rational maps almost entirely in future.

Proposition 7.4.18

If W is an irreducible subset of $M_n(F)$, with F infinite, then so too is the set

$$X = \bigcup_{T \in GL_n(F)} T^{-1}WT$$

of all matrices that are similar to some matrix in W . In particular, every similarity class of matrices is an irreducible subset.

Proof

The rational map

$$\theta : GL_n(F) \times W \longrightarrow X$$

$$(T, A) \longmapsto T^{-1}AT = (1/\det T)(\operatorname{adj} T)AT$$

(see Example 7.1.12) is defined everywhere on its domain. Hence, θ is a Zariski continuous mapping onto X by Proposition 7.4.17. By Remark 7.4.5, we know $GL_n(F)$ is irreducible. Therefore by Proposition 7.4.16, the domain of θ is irreducible because it is a product of irreducible sets. Now X is irreducible by Proposition 7.4.14 because it is the image of an irreducible set under a Zariski continuous mapping. The final statement of the proposition follows from the fact that singleton subsets W of $M_n(F)$ are irreducible. \square

Let us illustrate some of the ideas and results of this section.

Example 7.4.19

We wish to show that, over an algebraically closed field F , the special linear group $SL_n(F)$ is an irreducible subvariety of $M_n(F)$ (where the latter is identified with \mathbb{A}^{n^2}). As we observed in Example 7.1.9, $SL_n(F)$ is the hypersurface determined by the polynomial $f(x_{11}, x_{12}, \dots, x_{nn}) = \det(x_{ij}) - 1$. Irreducibility is therefore equivalent to f being a power of an irreducible polynomial (as seen

in Example 7.4.12). However, that would be difficult to establish directly. We take a different approach using known irreducible varieties, products, and Zariski continuous maps.

Let

$$W = \left\{ \left[\begin{array}{c|c} A & * \\ \hline 0 & b \end{array} \right] \in SL_n(F) : A \in M_{n-1}(F), b \in F, b \det A = 1 \right\}.$$

Then W is a subvariety of $M_n(F)$ because the matrices (b_{ij}) in W are determined by the polynomial equations $b_{n1} = b_{n2} = \cdots = b_{n,n-1} = 0$ and $\det(b_{ij}) - 1 = 0$. Consider the rational map

$$\theta : GL_{n-1}(F) \times \mathbb{A}^{n-1} \longrightarrow W$$

$$(A, a_1, a_2, \dots, a_{n-1}) \longmapsto \left[\begin{array}{c|c} A & \begin{matrix} a_1 \\ \vdots \\ a_{n-1} \end{matrix} \\ \hline 0 & 1/\det A \end{array} \right].$$

Note that this rational map is defined everywhere on its domain, so by Proposition 7.4.17, θ is Zariski continuous. As in Remark 7.4.5, we know $GL_{n-1}(F)$ is irreducible because it is a nonempty open subset of the irreducible variety $M_{n-1}(F)$. By Proposition 7.4.2, \mathbb{A}^{n-1} is irreducible. Therefore the domain of θ is irreducible by Proposition 7.4.16. Consequently Proposition 7.4.14 shows that W (the image of θ) is an irreducible subset of $M_n(F)$. Finally, $SL_n(F)$ is irreducible by Proposition 7.4.18 because, over an algebraically closed field, every matrix is similar to an upper triangular matrix, whence

$$SL_n(F) = \bigcup_{T \in GL_n(F)} T^{-1}WT.$$

Example 7.4.20

The set S of $n \times n$ singular matrices is an irreducible subvariety of $M_n(F)$ for an infinite field F . For

$$S = \bigcup_{T \in GL_n(F)} T^{-1}WT$$

where W is the set of $n \times n$ matrices that have a zero first column. Note that W is irreducible because it is naturally affine isomorphic to \mathbb{A}^{n^2-n} . Therefore, S is irreducible by Proposition 7.4.18.

This gives us a quick way of seeing that the set

$$X = \{A \in M_2(F) : \text{rank } A = 1\}$$

is not Zariski closed in $M_2(F)$, because otherwise its union with $\{0\}$ would contradict the irreducibility of the set of 2×2 singular matrices. Neither is X a Zariski open subset, otherwise its complement would be closed but contain the dense subset $GL_2(F)$ of $M_2(F)$, forcing the complement to be all of $M_2(F)$.¹¹ \square

Test Question. Is the variety of all $n \times n$ nilpotent matrices irreducible? (The answer is given during the proof of Lemma 7.10.3.)

7.5 EQUIVALENCE OF ASD FOR MATRICES AND IRREDUCIBILITY OF $\mathcal{C}(k, n)$

In Chapter 6, we studied the question of when k commuting $n \times n$ matrices A_1, A_2, \dots, A_k over the complex field \mathbb{C} can be approximated by simultaneously diagonalizable matrices B_1, B_2, \dots, B_k . That is, given any $\epsilon > 0$, the B_i can be chosen to be simultaneously diagonalizable and with $\|B_i - A_i\| < \epsilon$ for all i . We referred to this as the *ASD property*. The ASD property for all k commuting $n \times n$ complex matrices can be expressed as a denseness property relative to the usual Euclidean complex topology (the metric topology) in \mathbb{C}^{kn^2} . Namely, the set $\mathcal{D}(k, n)$ of all k -tuples (B_1, B_2, \dots, B_k) of simultaneously diagonalizable complex $n \times n$ matrices is a Euclidean dense subset of the set $\mathcal{C}(k, n)$ of all k -tuples (A_1, A_2, \dots, A_k) of commuting $n \times n$ complex matrices. Now $\mathcal{C}(k, n)$ is a natural subvariety of \mathbb{A}^{kn^2} (over any field; see Example 7.1.5) and so density properties with respect to the Zariski topology crop up naturally. For instance, if the variety is irreducible, then every nonempty Zariski open subset is Zariski dense. It may seem a long shot, given how much weaker the Zariski topology is compared with the Euclidean topology, but must nonempty Zariski open subsets U of an irreducible complex affine variety V be actually Euclidean dense in V ? The answer is surprisingly “yes”! What’s more, this enables us to show that the ASD property for k commuting complex matrices is equivalent to the irreducibility of the affine complex variety $\mathcal{C}(k, n)$. In this section we shall establish that equivalence, and in later sections we shall explore its implications.

11. The reader can compare this argument with his or her own argument at the end of Example 7.2.3.

Recall that in the Euclidean topology on \mathbb{C}^m , a basis of open subsets is supplied by the open balls

$$B(a, r) = \{b = (b_1, b_2, \dots, b_m) \in \mathbb{C}^m : \|b - a\| < r\}$$

where $a \in \mathbb{C}^m$ and $r \in \mathbb{R}^+$. Thus, for subsets $B \subseteq A \subseteq \mathbb{C}^m$, B is Euclidean dense in A if and only if for each $a \in A$ and $\epsilon > 0$, we always have $B(a, \epsilon) \cap B \neq \emptyset$, that is, $\|b - a\| < \epsilon$ for some $b \in B$. Another way of saying this is that for each member $a = (a_1, a_2, \dots, a_m) \in A$, we can ϵ -perturb each a_i to some b_i for $i = 1, 2, \dots, m$ such that $(b_1, b_2, \dots, b_m) \in B$.

We now give the important connection between Zariski denseness and Euclidean denseness. It deserves a name and, for want of a better one, we call it the Denseness Theorem. The result is certainly not well known. There are proofs in the literature, dating back at least to Mumford's 1966 text *Lectures on Curves on an Algebraic Surface* (see also Mumford's 1999 "*Red Book*," I 10, Corollary 1 on p. 60), but they require a very good knowledge of algebraic geometry, complex manifolds, and their ilk. They are not suitable for a book such as ours. It is perhaps not surprising that there are no really elementary proofs, because the corresponding statement is false for irreducible *real* affine varieties (see Example 7.11.2). Thankfully, S. Paul Smith of the University of Washington has provided us with a very nice proof for the theorem, which is pretty accessible to the nonexpert.¹² We present it in the last section for those who are interested.

Theorem 7.5.1 (The Denseness Theorem)

Nonempty Zariski open subsets of an irreducible complex affine variety V are Euclidean dense in V .

Here now is our *raison d'être* for enlisting algebraic geometry in our approximately simultaneously diagonalizable (ASD) matrix problem:

Theorem 7.5.2

All collections of k commuting $n \times n$ complex matrices are approximately simultaneously diagonalizable if and only if the variety $\hat{\mathcal{C}}(k, n)$ of k -tuples of commuting $n \times n$ complex matrices is irreducible.

Proof

Throughout, we work with the field $F = \mathbb{C}$ of complex numbers. First, assume $\mathcal{C}(k, n)$ is irreducible for some given k and n . We wish to establish ASD for any k commuting $n \times n$ complex matrices. We identify $\mathcal{C}(k, n)$ with a subset of \mathbb{A}^{kn^2} in

12. Keith Conrad, Ken Goodearl, and Ross Moore also made helpful comments on the proof. We thank Herb Clemens for first confirming to us the existence of the Denseness Theorem.

the standard way. Let

$$U = \{(B_1, B_2, \dots, B_k) \in \mathcal{C}(k, n) : B_1 \text{ has } n \text{ distinct eigenvalues}\}.$$

Claim: U is a Zariski open subset of $\mathcal{C}(k, n)$.

For if the eigenvalues of B_1 are $\lambda_1, \lambda_2, \dots, \lambda_n$, distinctness of these is determined by the nonvanishing of the polynomial

$$p(\lambda_1, \lambda_2, \dots, \lambda_n) = \prod_{i \neq j} (\lambda_i - \lambda_j).$$

This polynomial is symmetric in $\lambda_1, \lambda_2, \dots, \lambda_n$, hence, by Proposition 7.1.10, can be expressed as a polynomial in the entries of B_1 . Thus, there is a polynomial $f(x_{11}, x_{12}, \dots, x_{nn})$ in n^2 variables whose nonvanishing set in $M_n(F)$ is the open subset of all matrices that have distinct eigenvalues. The nonvanishing set of the corresponding polynomial (viewed as having kn^2 variables, but with zero coefficients for the added variables) in $\mathcal{C}(k, n)$ is therefore U . Thus, U is Zariski open.

Now fix $(A_1, A_2, \dots, A_k) \in \mathcal{C}(k, n)$. Since U is a nonempty open subset of the irreducible variety $\mathcal{C}(k, n)$, by the Denseness Theorem 7.5.1, U is Euclidean dense in $\mathcal{C}(k, n)$. Hence, there are ϵ -perturbations B_i of A_i for $i = 1, 2, \dots, k$ such that $(B_1, B_2, \dots, B_k) \in U$. Since B_1 is 1-regular, its centralizer $\mathcal{C}(B_1)$ is $F[B_1]$ by Proposition 1.1.2, so $B_i = p_i(B_1)$ for some polynomials $p_i(x) \in F[x]$ for $i = 2, 3, \dots, k$. Now B_1 is certainly diagonalizable because it has distinct eigenvalues. But if $C \in GL_n(F)$ diagonalizes B_1 , then C also diagonalizes the other B_i because they are polynomials in B_1 . Thus, we have perturbed A_1, A_2, \dots, A_k to simultaneously diagonalizable B_1, B_2, \dots, B_k , as desired.

Conversely, suppose for some given k and n that the ASD property holds for all k commuting $n \times n$ complex matrices. Let $V = \mathcal{C}(k, n) \subseteq \mathbb{A}^{kn^2}$. We wish to show that the variety V is irreducible. Let

$$W = \{(D_1, D_2, \dots, D_k) \in V : \text{each } D_i \text{ is diagonal}\}.$$

(We really do mean D_i diagonal here, not just diagonalizable.) Clearly W is a subvariety (determined by the polynomials x_{ij} for $i \neq j$ and over the appropriate range) and W is naturally isomorphic to \mathbb{A}^{kn} . In particular, W is irreducible by Proposition 7.4.2. Note that the ASD property implies (in fact is equivalent to) the set

$$X = \bigcup_{T \in GL_n(F)} T^{-1}WT$$

of k -tuples (B_1, B_2, \dots, B_k) of simultaneously diagonalizable $n \times n$ matrices being Euclidean dense in V . In particular, X is Zariski dense in V . But

X is irreducible by the obvious extension of Proposition 7.4.18, whence its Zariski closure V is also irreducible by Proposition 7.4.3. Thus, $\mathcal{C}(k, n)$ is irreducible. \square

7.6 GERSTENHABER REVISITED

In the mid-1950s, Motzkin and Taussky showed that the variety $\mathcal{C}(2, n)$ of commuting pairs of matrices over an algebraically closed field F is irreducible. In this section, we establish the Motzkin–Taussky theorem using a short proof of Guralnick. As a corollary, we show how one can quickly deduce Gerstenhaber’s theorem using an elegant argument given by Guralnick in the early 1990s. This is a very good illustration of the power of algebraic geometry arguments in even purely algebraic settings. Note that Gerstenhaber’s theorem, which we studied in Chapter 5, seemingly has nothing to do with algebraic geometry!

Theorem 7.6.1 (Motzkin–Taussky)

Over an algebraically closed field F , the variety $\mathcal{C}(2, n)$ of commuting pairs of $n \times n$ matrices is irreducible for all $n \geq 1$.

Proof

First, we observe that every $A \in M_n(F)$ commutes with some 1-regular matrix R . (Note, however, that A may not commute with some matrix having distinct eigenvalues unless A is diagonalizable.) For we can suppose A is in Jordan form, say $A = \text{diag}(J_1, J_2, \dots, J_r)$ where J_i is a basic Jordan block with eigenvalue b_i . Then we can take the 1-regular matrix

$$R = \text{diag}((a_1 - b_1)I + J_1, (a_2 - b_2)I + J_2, \dots, (a_r - b_r)I + J_r)$$

for any distinct a_1, a_2, \dots, a_r (note that R has a_1, a_2, \dots, a_r as its eigenvalues, and they each have geometric multiplicity 1), and clearly R commutes with A . (Here is an instance where the Weyr form does not work as well as the Jordan form.) Let $V = \mathcal{C}(2, n)$ and let

$$U = \{(A, R) \in V : R \text{ is 1-regular}\}.$$

Claim: U is Zariski dense in V .

Since 1-regularity is an open condition (Example 7.2.4), U is a Zariski open subset, so it has to be Zariski dense if V is irreducible. That won’t help us at this point, however, because we don’t yet know that V is irreducible. Let $(A, B) \in V$ be given, and let

$$U(f) = \{(X, Y) \in M_n(F) \times M_n(F) : f(X, Y) \neq 0\}$$

be any open basis set of $M_n(F) \times M_n(F) (\cong \mathbb{A}^{2n^2})$ containing (A, B) . That is, f is some polynomial in $2n^2$ variables that does not vanish at (A, B) . We shall show that $U(f) \cap U$ is nonempty. Choose a 1-regular matrix R that commutes with A . Then, since 1-regularity is an open condition determined by the nonvanishing of one of a finite number of polynomials, there is a polynomial p that doesn't vanish at R , and its nonvanishing at any other matrix $X \in M_n(F)$ is *sufficient* (but not necessary) for X to be 1-regular. Since p doesn't vanish at R , the single variable polynomial $p(xB + R) \in F[x]$ must be *nonzero*.¹³ Therefore, since nonzero polynomials in a single variable have only finitely many roots, $p(bB + R)$ is nonzero for all but a finite number of $b \in F$. Therefore, $B + cR = c(c^{-1}B + R)$ is 1-regular for all but a finite number of nonzero $c \in F$. Moreover, since $f(A, B) \neq 0$, we similarly have $f(A, B + cR) \neq 0$ for almost all $c \in F$. Hence (since F is infinite), we can find $c \in F$ such that both $f(A, B + cR) \neq 0$ and $B + cR$ is 1-regular. Inasmuch as $B + cR$ commutes with A , this shows $(A, B + cR) \in U(f) \cap U$. Thus, (A, B) is in the Zariski closure of U , which establishes our claim.

Next consider the polynomial map

$$f : F^n \times M_n(F) \longrightarrow V$$

$$(b_0, b_1, \dots, b_{n-1}, B) \longmapsto (b_0I + b_1B + \dots + b_{n-1}B^{n-1}, B).$$

The domain of f is irreducible, by Proposition 7.4.16, because it is a product of irreducible sets, and f is Zariski continuous by Proposition 7.2.7, whence the image Z of f is irreducible by Proposition 7.4.14. Consequently, the closure \bar{Z} of Z in V is also irreducible (by Proposition 7.4.3). But Z contains the subset U (because $\mathcal{C}(R) = F[R]$ for 1-regular matrices R by Proposition 3.2.4), and U is dense in V , so $\bar{Z} = \bar{U} = V$. Thus, V is irreducible. \square

An immediate corollary to this Motzkin–Tausky theorem is the earlier Motzkin–Tausky Theorem 6.8.1 that we established in Chapter 6.

Corollary 7.6.2

All pairs of commuting $n \times n$ complex matrices can be approximately simultaneously diagonalized.

Proof

Theorems 7.5.2 and 7.6.1. \square

Corollary 7.6.3 (Gerstenhaber's Theorem)

For any pair of commuting matrices $A, B \in M_n(F)$, the dimension of the subalgebra $F[A, B]$ generated by A and B can't exceed n .

13. An important part of the argument, yet easily overlooked.

Proof

As noted in our original proof (see the introduction to Chapter 5), there is no loss of generality in assuming F is algebraically closed. Let $V = \mathcal{C}(2, n)$ and let $(A, B) \in V$. Form the $n^2 \times n^2$ matrix M whose columns are the matrices $A^i B^j$ written as $n^2 \times 1$ column vectors, for $0 \leq i, j \leq n-1$. For example, for the pair of commuting 2×2 matrices A, B given by

$$A = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 4 \\ -2 & 7 \end{bmatrix}, \quad \text{with } AB = BA = \begin{bmatrix} -1 & 18 \\ -9 & 17 \end{bmatrix},$$

if we list the $A^i B^j$ in the particular order $A^0 B^0, A^1 B^0, A^0 B^1, A^1 B^1$, then

$$M = \begin{bmatrix} 1 & 1 & 3 & -1 \\ 0 & 2 & 4 & 18 \\ 0 & -1 & -2 & -9 \\ 1 & 3 & 7 & 17 \end{bmatrix}.$$

By the Cayley–Hamilton theorem, $F[A, B]$ is spanned as a vector space by the n^2 matrices $A^i B^j$ for $0 \leq i, j \leq n-1$, and so we have

$$(*) \quad \dim F[A, B] \leq n \iff \text{rank } M \leq n.$$

(Our 4×4 example M has rank 2 so $\dim F[A, B] = 2$, in agreement with the theorem.¹⁴) The $(*)$ condition is some polynomial condition $p_1(A, B) = p_2(A, B) = \cdots = p_k(A, B) = 0$ in the entries of A and B determined by some finite collection of polynomials p_1, p_2, \dots, p_k in $2n^2$ variables, because $(*)$ just says all $(n+1) \times (n+1)$ submatrices of M have zero determinant. Thus, the set

$$W = \{(A, B) \in V : \dim F[A, B] \leq n\}$$

is a Zariski closed subset of V . But W contains the nonempty Zariski open subset

$$U = \{(A, R) \in V : R \text{ is 1-regular}\}$$

because, for a 1-regular matrix R , any centralizing A must be in $F[R]$, whence $\dim F[A, R] = \dim F[R] = n$. (See Proposition 3.2.4.) By irreducibility of V , its nonempty open subset U must be dense in V . Thus, W is also dense in V and so, since W is closed as well, we must have $W = V$. Hence, $\dim F[A, B] \leq n$ for all commuting A and B . \square

14. To our relief!

Remark 7.6.4

In the case of $n \times n$ commuting nilpotent matrices A and B , we gave in Theorem 5.3.1 an explicit spanning set \mathcal{B} for $F[A, B]$, whose n members were certain matrices of the form $A^i B^j$. It is not clear how one could derive this result from algebraic geometry, using the irreducibility of $\mathcal{C}(2, n)$. \square

Before looking at further corollaries to the Motzkin–Tausky theorem, we record two useful necessary conditions for the variety $\mathcal{C}(k, n)$ to be irreducible.

Proposition 7.6.5

Let F be an algebraically closed field. If $\mathcal{C}(k, n)$ is irreducible (for some given k and n), then for all k commuting $n \times n$ matrices A_1, \dots, A_k we have:

- (1) $\dim F[A_1, \dots, A_k] \leq n$
- (2) $\dim \mathcal{C}(A_1, \dots, A_k) \geq n$.

Proof

Let

$$V = \mathcal{C}(k, n),$$

$$W_1 = \{(A_1, \dots, A_k) \in V : \dim F[A_1, \dots, A_k] \leq n\},$$

$$W_2 = \{(A_1, \dots, A_k) \in V : \dim \mathcal{C}(A_1, \dots, A_k) \geq n\}.$$

Observe that, in view of Proposition 3.2.4, both W_1 and W_2 contain the nonempty Zariski open subset

$$U = \{(A_1, \dots, A_k) \in V : A_1 \text{ is 1-regular}\}.$$

Therefore, it suffices to show that W_1 and W_2 are Zariski closed, for then irreducibility of V implies $\overline{U} = V$ and thus $W_1 = \overline{W_1} \supseteq \overline{U} = V$, giving $W_1 = V$. And similarly $W_2 = V$.

That W_1 is closed follows from the same sort of argument used in the proof of Corollary 7.6.3 when $k = 2$. We now show W_2 is Zariski closed by an argument similar to that used in Lemma 6.5.3. Fix $(A_1, \dots, A_k) \in V$ and consider the map $\theta : X \mapsto ([A_1, X], \dots, [A_k, X])$ of $M_n(F)$ into the product $M_n(F) \times \dots \times M_n(F)$ of k copies of $M_n(F)$. (Here, $[P, Q] = PQ - QP$ is the additive commutator.) Then θ is a linear transformation whose kernel is $\mathcal{C}(A_1, \dots, A_k)$. Hence, the condition for (A_1, \dots, A_k) to be in W_2 is that $\text{nullity}(\theta) \geq n$, equivalently

$$(*) \quad \text{rank}(\theta) \leq n^2 - n$$

(because $\text{rank}(\theta) + \text{nullity}(\theta) = \dim M_n(F) = n^2$). In terms of the $kn^2 \times n^2$ matrix M of θ relative to the pair of standard bases for $M_n(F)$ and $M_n(F) \times \dots \times M_n(F)$, we see that $(*)$ can be expressed as the vanishing of the determinants

of the $(n^2 - n + 1) \times (n^2 - n + 1)$ submatrices of M . In turn these are polynomial equations in the entries of A_1, \dots, A_k . Hence, W_2 is indeed closed, as asserted. \square

Here now are two further corollaries to the Motzkin–Tausky theorem. The first is one half of the Laffey–Lazarus and Neubauer–Saltman Theorem 5.4.4 that we discussed in Chapter 5:

Corollary 7.6.6

Let A and B be commuting $n \times n$ matrices over an algebraically closed field F , and suppose $F[A, B]$ is a maximal commutative subalgebra of $M_n(F)$. Then $\dim F[A, B] = n$.

Proof

By Corollary 7.6.3 (Gerstenhaber’s theorem), $\dim F[A, B] \leq n$. We have $\mathcal{C}(A, B) = F[A, B]$ here because $F[A, B]$ is a maximal commutative subalgebra of $M_n(F)$. Therefore, by the Motzkin–Tausky Theorem 7.6.1 and Proposition 7.6.5 (2), we must have $\dim F[A, B] = \dim \mathcal{C}(A, B) \geq n$. Thus, $\dim F[A, B] = n$. \square

Corollary 7.6.7

For $n \leq 3$, the variety $\mathcal{C}(k, n)$ is irreducible for all k .

Proof

Since $n \leq 3$, every commutative subalgebra \mathcal{A} of $M_n(F)$ is (at most) 2-generated. One can deduce this as in Example 5.4.5, or check it out directly. Therefore, there exist $A, B \in \mathcal{A}$ such that every member C of \mathcal{A} is in the span of the monomials $A^i B^j$ for $i, j = 0, 1, \dots, n-1$, that is $C = f(A, B)$ for some polynomial $f(x, y)$ in the set

$$P = \{f(x, y) \in F[x, y] : f \text{ has degree at most } n-1 \text{ in both } x \text{ and } y\}.$$

The Zariski continuous map

$$\begin{aligned} \eta : \mathcal{C}(2, n) \times P^k &\longrightarrow \mathcal{C}(k, n) \\ (A, B, f_1, f_2, \dots, f_k) &\longmapsto (f_1(A, B), f_2(A, B), \dots, f_k(A, B)) \end{aligned}$$

is therefore onto. Moreover, by the Motzkin–Tausky theorem, the domain of η is irreducible because it is a product of the irreducible varieties $\mathcal{C}(2, n)$ and $P \cong \mathbb{A}^{n^2}$. Therefore, its image $\mathcal{C}(k, n)$ is irreducible by Proposition 7.4.14. \square

We conclude this section by showing that, in contrast to the Motzkin–Tausky result and the previous corollary, the varieties $\mathcal{C}(k, n)$ are never irreducible when $k \geq 4$ and $n \geq 4$.

Proposition 7.6.8

Over an algebraically closed field F , the variety $\mathcal{C}(k, n)$ of k -tuples of commuting $n \times n$ matrices is reducible when $k \geq 4$ and $n \geq 4$.

Proof

For these values of k and n we know from the argument in Example 6.3.4 that there are k (in fact 4) commuting $n \times n$ matrices that generate a subalgebra of $M_n(F)$ of dimension greater than n . Therefore, by Proposition 7.6.5 (1), $\mathcal{C}(k, n)$ can't be irreducible. \square

That now leaves only the question of when $\mathcal{C}(3, n)$ is irreducible. However, this is considerably trickier and, as we type this, still partly open. Armed with the Weyr form, we take up the challenge in Sections 7.9 and 7.10.

7.7 CO-ORDINATE RINGS OF VARIETIES

A good way of introducing the proven tools of commutative ring theory into the study of varieties is through co-ordinate rings. The **co-ordinate ring** of a nonempty affine variety $V \subseteq \mathbb{A}^n$ is the ring of all polynomial functions $f : V \rightarrow \mathbb{A}^1$. (Addition and multiplication in the ring are pointwise.) Although we will follow the modern convention of denoting the co-ordinate ring by $F[V]$, we must confess to some misgivings about this notation, because it somehow suggests the ring elements are “polynomial expressions in V .”¹⁵ We can make an identical definition of the co-ordinate ring $F[X]$ of an arbitrary subset X of \mathbb{A}^n . By Proposition 7.4.7, $F[X]$ is naturally isomorphic to the co-ordinate ring of the variety \overline{X} , where \overline{X} is the Zariski closure of X .

An Aside. Before we get into some deeper aspects of algebraic geometry, this may be a good place to clear up any possible misconceptions about the nature of polynomials, evaluating polynomials, polynomial maps, etc. These notions are at the heart of algebraic geometry. Among the “pure-breds” of their respective disciplines, an algebraist and an analyst may have a different take on the role of the “indeterminate” x in a polynomial $p(x)$ over, say, the reals \mathbb{R} . The algebraist may well say that x is a concrete object in the (associative) algebra $\mathcal{A} = \mathbb{R}[x]$ which has $\{1, x, x^2, \dots, x^n, \dots\}$ as a vector space basis. (After that, the way in

15. $F[V]$ also conflicts with other uses of this notation, particularly for the subalgebra $F[A]$ generated by a matrix A .

which polynomials behave is completely determined, including their addition and multiplication.) The algebra \mathcal{A} is akin to the usual ring \mathbb{Z} of integers. It is a unique factorization domain (it even has a Euclidean algorithm), and so on. To the algebraist, x is just as real an algebraic object as the prime number 3. There is nothing “airy-fairy” about x , and the term “indeterminate” is a misnomer. The analyst, on the other hand, may say that x is just the variable in a function

$$p : \mathbb{R} \longrightarrow \mathbb{R}$$

that takes the form $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ for some fixed $a_i \in \mathbb{R}$. These continuous functions are beautifully behaved, easy to differentiate and integrate, and can be used to describe many other useful functions, through the use of Taylor polynomials, Taylor series, and so on. To the analyst, the term “indeterminate” is quite appropriate because x represents a general real number.¹⁶

Both views have merits. Moreover, a polynomial $p(x) \in \mathbb{R}[x]$ in the algebraist’s sense is faithfully represented by the corresponding polynomial function in the analyst’s sense; in fact, the correspondence is an isomorphism with respect to the standard operations. But when it comes to general fields F and polynomial maps $p : X \rightarrow F$ defined on some subset X of \mathbb{A}^n , one must be careful to distinguish between the polynomial $p(x_1, \dots, x_n)$ as an algebraic object in the algebra $F[x_1, \dots, x_n]$ of polynomials and the associated polynomial mapping p , which is a function from X into the base field F . These are mathematically quite different. Here one can view $\mathcal{A} = F[x_1, \dots, x_n]$ as a commutative, associative algebra over F in which the monomials

$$x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \quad (a_i \text{ nonnegative integers})$$

are distinct for distinct choices of the exponents a_i , and these monomials form a vector space basis for \mathcal{A} .¹⁷

“Evaluating” a polynomial $p(x_1, \dots, x_n) \in F[x_1, \dots, x_n]$ in more than one variable, at a particular point $(a_1, \dots, a_n) \in \mathbb{A}^n$, is the natural extension of the single variable case: for each $i = 1, \dots, n$, replace each occurrence of x_i in p by a_i , and compute the resulting linear combination of the various products according to the arithmetic in the particular field F . We denote this value by $p(a_1, \dots, a_n)$. One point to watch, however. What if some x_i does not appear

16. At this point, should scuffling break out, call a geometer to restore order.

17. The sophisticated statement of this is that $F[x_1, \dots, x_n]$ is the semigroup algebra over F of the free commutative semigroup generated by x_1, \dots, x_n .

in p with a nonzero coefficient? Then just ignore a_i in the substitution process (alternatively, include x_i in p with a zero coefficient). For instance, if

$$p(x_1, x_2, x_3) = 1 + x_1x_3 + x_3^2,$$

then $p(2, 4, 5) = 1 + (2 \times 5) + 5^2 = 36$.

Given a subset X of \mathbb{A}^n , we can associate with each polynomial $p(x_1, \dots, x_n) \in F[x_1, \dots, x_n]$ the polynomial map (often denoted by the same letter p)

$$f : X \longrightarrow \mathbb{A}^1$$

$$f(a_1, \dots, a_n) = p(a_1, \dots, a_n) \text{ for all } (a_1, \dots, a_n) \in X.$$

But two different polynomials may well induce the same polynomial map, whence the need to distinguish, for example, between when f is the zero mapping and when p is the zero polynomial (that is, p is the zero of the algebra \mathcal{A}). For instance, the polynomial $p(x, y) = -2 - x + 2x^2 + 2y^2 + xy^2 + x^3 = (x + 2)(x^2 + y^2 - 1) \in \mathbb{R}[x, y]$ is not the zero polynomial but its associated polynomial mapping

$$f : X \longrightarrow \mathbb{R}$$

$$f(a_1, a_2) = -2 - a_1 + 2a_1^2 + 2a_2^2 + a_1a_2^2 + a_1^3$$

on the unit circle $X = \{(x, y) : x^2 + y^2 = 1\}$ is the zero mapping. \square

We now return to the main chase. The following is a most useful result. Recall that in Section 7.4, we associated to each subvariety V of \mathbb{A}^n the ideal $I(V)$ of all polynomial functions that vanish on V .

Proposition 7.7.1

Let V be a nonempty affine subvariety of \mathbb{A}^n . Then:

- (1) $F[V] \cong F[x_1, x_2, \dots, x_n]/I(V)$ as F -algebras.
- (2) V is an irreducible subvariety precisely when $F[V]$ is an integral domain, equivalently, $I(V)$ is a prime ideal of $F[x_1, x_2, \dots, x_n]$.

Proof

The map

$$\theta : F[x_1, x_2, \dots, x_n] \rightarrow F[V], f \mapsto \text{polynomial function of } V \text{ induced by } f$$

is an onto algebra homomorphism whose kernel is $I(V)$. Thus, (1) follows from the fundamental homomorphism theorem. In our new language, Proposition 7.4.8 says that V is irreducible if and only if $F[V]$ is an integral domain (has no zero

divisors). Recall that by definition, an ideal P of a commutative ring R is *prime* if P is proper ($P \neq R$) and if $ab \in P$ for $a, b \in R$ implies either $a \in P$ or $b \in P$. Hence, (2) follows from (1) because an ideal I of a commutative ring R is prime exactly when the factor ring R/I is an integral domain. \square

This is an appropriate place to introduce (in the next theorem) another version of Hilbert's Nullstellensatz 7.3.3, in terms of our correspondence $J \mapsto V(J)$ between the ideals J of $F[x_1, x_2, \dots, x_n]$ and subvarieties of \mathbb{A}^n , and the reverse correspondence $W \mapsto I(W)$. By definition, the **radical** of an ideal J of a commutative ring R is the ideal

$$\sqrt{J} = \{a \in R : a^k \in J \text{ for some positive integer } k\}.$$

Thus, \sqrt{J}/J is the ideal of all nilpotent elements of the factor ring R/J . If $\sqrt{J} = J$, then J is called a **radical ideal**. Thus, radical ideals are those for which the factor ring has no nonzero nilpotents. In particular, every prime ideal is radical.

Theorem 7.7.2

Let F be an algebraically closed field and let J be an ideal of $F[x_1, x_2, \dots, x_n]$. Then $I(V(J)) = \sqrt{J}$. In particular, two ideals J_1 and J_2 determine the same subvariety if and only if they have the same radical.

Proof

The containment $\sqrt{J} \subseteq I(V(J))$ is the easy part. The reverse containment is usually proved using the “Rabinowitsch’s trick.” We refer the interested reader to Chapter 1 of Hulek’s text *Elementary Algebraic Geometry* or Section 4.5 of Eisenbud’s text *Commutative Algebra with a View Toward Algebraic Geometry*.

If J_1 and J_2 determine the same subvariety, then $\sqrt{J_1} = I(V(J_1)) = I(V(J_2)) = \sqrt{J_2}$, whence J_1 and J_2 have the same radical. Conversely, suppose $\sqrt{J_1} = \sqrt{J_2}$. By Proposition 7.4.9(4), we have $V(J_1) = V(I(V(J_1))) = V(\sqrt{J_1}) = V(\sqrt{J_2}) = V(I(V(J_2))) = V(J_2)$. Thus, J_1 and J_2 determine the same subvariety. \square

It is well known that the ring $F[x]$ of polynomials in a single indeterminate over a field F is a unique factorization domain (UFD). The same applies to the polynomial ring $F[x_1, x_2, \dots, x_n]$ in a finite number of indeterminates, because if U is a UFD, then so too is the polynomial ring $U[x]$. (See Jacobson’s *Basic Algebra I*, Theorem 2.25.) Given a hypersurface $V(f)$ of \mathbb{A}^n , therefore, it is natural to inquire how the unique irreducible factorization $f = f_1^{k_1} f_2^{k_2} \cdots f_m^{k_m}$ of f relates to the irreducible decomposition of $V(f)$. This is revealed in the following corollary.

Corollary 7.7.3

Let $R = F[x_1, x_2, \dots, x_n]$ where F is an algebraically closed field.

- (1) $I(V(J)) = J$ for every radical ideal (in particular, every prime ideal) J of R .
- (2) For any hypersurface $V = V(f)$ of \mathbb{A}^n determined by a nonconstant polynomial $f \in R$ with irreducible factorization $f = f_1^{k_1} f_2^{k_2} \cdots f_m^{k_m}$, the ideal $I(V)$ is generated by $f_1 f_2 \cdots f_m$.
- (3) For any nonconstant polynomial $f \in R$, the hypersurface $V = V(f)$ is an irreducible variety if and only if f is a power of an irreducible polynomial.
- (4) For a general polynomial f as in (2), the $V(f_i)$ are the irreducible components of $V(f)$.

Proof

(1) is immediate from Theorem 7.7.2. For (2), $V = V(J)$ where J is the principal ideal fR and so $I(V) = \sqrt{J} = f_1 f_2 \cdots f_m R$. Now recall from Proposition 7.7.1 (2) that V is irreducible if and only if $I(V)$ is a prime ideal. From (2) this occurs precisely when $m = 1$, establishing (3).

To see (4), note first that

$$V(f) = V(f_1^{k_1}) \cup V(f_2^{k_2}) \cup \cdots \cup V(f_m^{k_m}) = V(f_1) \cup V(f_2) \cup \cdots \cup V(f_m)$$

gives a decomposition into irreducible varieties by (3). On the other hand, the decomposition is irredundant because if $V(f_i) \subseteq V(f_j)$, then by (1) we have $f_j R = I(V(f_j)) \subseteq I(V(f_i)) = f_i R$, whence $i = j$ because f_i and f_j are irreducible. \square

Theorem 7.7.2 and Proposition 7.4.9 (4) now tell us that the V and I maps induce inverse bijections

$$\begin{array}{ccc} & V & \\ \{\text{radical ideals of } F[x_1, x_2, \dots, x_n]\} & \xrightarrow{\quad} & \{\text{subvarieties of } \mathbb{A}^n\} \\ & \xleftarrow{\quad} & \\ & I & \end{array}$$

under which the prime ideals correspond to (nonempty) irreducible subvarieties (and maximal ideals correspond to points in affine n -space).

Our particular interest in co-ordinate rings really relates only to the “dimension” of a variety (in the sense of algebraic geometry), which we shall discuss in the next section. There is much more that can be said about co-ordinate rings, even at an elementary level. For the reader’s edification, we document some properties in the following remarks (it is assumed our base field F is algebraically closed), even though we don’t intend making use of them. The proofs are not difficult. The interested reader can read the details in Hulek’s Chapter 1.

Remarks 7.7.4

- (1) The algebras that occur as co-ordinate rings of affine varieties are precisely those which are commutative, finitely generated, and have no nonzero nilpotent elements.
- (2) Let $V \subseteq \mathbb{A}^n$ and $W \subseteq \mathbb{A}^m$ be affine varieties. With each polynomial map $f : V \rightarrow W$, we associate an algebra homomorphism $f^* : F[W] \rightarrow F[V]$ by letting $f^*(g) = g \circ f$ for all $g \in F[W]$. Conversely, if $\phi : F[W] \rightarrow F[V]$ is an algebra homomorphism, there is a unique polynomial map $f : V \rightarrow W$ such that $\phi = f^*$.
- (3) As a consequence of (2), a variety is completely determined, to within isomorphism of varieties, by its co-ordinate ring: $V \cong W$ if and only if $F[V] \cong F[W]$.
- (4) We can view the affine varieties (over a fixed field F) as the objects in a category that has the polynomial maps as morphisms. The finitely generated commutative algebras (over F) that don't have nonzero nilpotents are likewise objects in a category in which the morphisms are the algebra homomorphisms. A neat way of viewing the $f \rightarrow f^*$ correspondence is to observe that the maps

$$\begin{aligned} V &\longmapsto F[V] \\ (f : V \rightarrow W) &\longmapsto (f^* : F[W] \rightarrow F[V]) \end{aligned}$$

induce a contravariant equivalence of the aforementioned categories. \square

Example 7.7.5

Suppose F is algebraically closed and consider the irreducible affine varieties V and W in \mathbb{A}^2 determined respectively by the polynomials $f(x, y) = y - x^2$ and $g(x, y) = y^2 - x^3$:

$$\begin{aligned} V &= \{(x, y) \in \mathbb{A}^2 : y = x^2\} \\ W &= \{(x, y) \in \mathbb{A}^2 : y^2 = x^3\}. \end{aligned}$$

Since f and g are irreducible polynomials, $I(V) = (f)$ (the principal ideal generated by f) and similarly $I(W) = (g)$, by Corollary 7.7.3 (2). Therefore by Proposition 7.7.1, the co-ordinate ring of V is

$$F[V] \cong F[x, y]/(y - x^2) \cong F[x],$$

and the co-ordinate ring of W is

$$F[W] \cong F[x, y]/(y^2 - x^3).$$

The two co-ordinate rings are not isomorphic because the first one is a unique factorization domain whereas the second is not. (Observe that $y^2 = x^3$ gives two

distinct irreducible factorizations in the second factor ring.) Hence, V and W are nonisomorphic varieties. However, they are Zariski homeomorphic because the maps $\eta : V \rightarrow W, \theta : W \rightarrow V$ given by

$$\eta(x, y) = (x^2, x^3), \quad \theta(x, y) = \begin{cases} (y/x, y^2/x^2) & \text{if } x \neq 0 \\ (0, 0) & \text{if } x = 0 \end{cases}$$

are mutually inverse Zariski continuous maps. On the other hand, since V and \mathbb{A}^1 have isomorphic co-ordinate rings (each isomorphic to $F[x]$), we can conclude that V is isomorphic to \mathbb{A}^1 . This is also easily seen directly, with $(x, y) \mapsto x$ and $x \mapsto (x, x^2)$ being inverse polynomial maps. \square

7.8 DIMENSION OF A VARIETY

Just as the dimension of a vector space plays a most fundamental role in the theory of vector spaces and their applications (for example to matrices), so too does the dimension of an affine variety play a fundamental role in algebraic geometry and its applications. The latter dimension, however, is a somewhat more sophisticated concept and requires more work to develop its basic properties. In this section we shall present the main properties we require of dimension for our particular applications to the matrix ASD question in Section 7.9.

Our standing assumption throughout this section is that F is an arbitrary field, since this suffices for the most basic properties of algebraic geometry dimension. On the other hand, deeper properties of dimension require F to be algebraically closed, an assumption signaled when appropriate.

If V is a vector subspace of F^n for an infinite field F , we will see that its vector space dimension and algebraic geometry dimension agree. Also as expected, in real affine 3-space, for example, it will turn out that points have dimension 0, curves have dimension 1, surfaces have dimension 2, and so on. Certainly \mathbb{A}^n will always have dimension n and its (nonempty) subvarieties will take dimension values in $\{0, 1, 2, \dots, n\}$. And if W is a subvariety of an irreducible variety V , then $\dim W \leq \dim V$ with strict inequality if $W \neq V$. There are a number of approaches to defining the algebraic geometry dimension of a variety. The quickest and most precise definition of the dimension of a subvariety V of \mathbb{A}^n is in terms of the co-ordinate rings of the irreducible components of V , and the notion of transcendence degree of a field extension. We will give this definition first and then develop a more informal approach to dimension that may be more suited to our readership. The essence of both approaches is measuring the “faithfulness” of representing polynomials $p \in F[x_1, x_2, \dots, x_n]$ by their corresponding polynomial map $p : V \rightarrow \mathbb{A}^1$. In short, just as

vector space dimension is expressible as the maximum degree of linear independence of vectors, algebraic geometry dimension of a subvariety $V \subseteq \mathbb{A}^n$ is expressible as the maximum degree of algebraic independence of polynomial functions on V .

We briefly recall some relevant concepts from field theory. Suppose $E \supseteq K$ is a field extension. An element $a \in E$ is **algebraic** over K if $f(a) = 0$ for some nonzero polynomial $f \in K[x]$. The unique monic polynomial $m(x) \in K[x]$ of least degree with $m(a) = 0$ is irreducible and is called the **minimal polynomial** of a over K . An element that is not algebraic is called **transcendental** over K . A finite subset $\{a_1, a_2, \dots, a_m\} \subseteq E$ is **algebraically dependent** over K if $f(a_1, a_2, \dots, a_m) = 0$ for some nonzero polynomial $f \in K[x_1, x_2, \dots, x_m]$. This is equivalent to some a_i being algebraic over the subfield generated by K and the other a_j . The set $\{a_1, a_2, \dots, a_m\}$ is thus **algebraically independent** over K precisely when the natural map

$$K[x_1, x_2, \dots, x_m] \rightarrow K[a_1, a_2, \dots, a_m], f(x_1, x_2, \dots, x_m) \mapsto f(a_1, a_2, \dots, a_m)$$

is a K -algebra isomorphism. Under this map, $x_i \mapsto a_i$, so when a_1, a_2, \dots, a_m are algebraically independent, they exhibit exactly the same degree of algebraic independence as do the indeterminates x_1, x_2, \dots, x_m . The subfield $K(a_1, a_2, \dots, a_m)$ is then naturally isomorphic to the field of rational functions $K(x_1, x_2, \dots, x_m)$ and is called a **purely transcendental extension** of K . An arbitrary subset S of E is **algebraically dependent** over K if some finite subset of S is such. A subset B of E is called a **transcendence base** of E over K if (1) B is algebraically independent over K and (2) every $a \in E$ is algebraic over the subfield $K(B)$ of E generated by K and B . Every field extension has a transcendence base, and any two such bases have the same cardinality. This common cardinality is denoted $\text{tr deg}_K E$ and is referred to as the **transcendence degree** of E over K . If E is finitely generated as a field over K , say, $E = K(a_1, a_2, \dots, a_s)$, then $\text{tr deg}_K E \leq s$. If $\text{tr deg}_K E = r$, then E can be viewed as having been obtained from K by first taking a purely transcendental extension $L = K(x_1, x_2, \dots, x_r)$ of K , and then following this by a finite (algebraic) extension of L . Chapter 8 of Jacobson's *Basic Algebra II* gives a nice treatment of these ideas. In particular, through his use of a “base” for a set X relative to an abstract “dependence relation,” Jacobson is able to give a unified treatment of vector space bases, transcendence bases, and more.

The key to the formal definition of dimension of an affine variety is this. If V is an irreducible variety, we know its co-ordinate ring $F[V]$ is an integral domain by Proposition 7.7.1. As such, it has a field of quotients E , which can be viewed as a field extension of F . We call E the **field of rational functions** of V and

denote it by $F(V)$. The transcendency degree of E over F is our candidate for the dimension of V .

Definition 7.8.1: The algebraic geometry **dimension** of a nonempty irreducible affine subvariety $V \subseteq \mathbb{A}^n$ is

$$\dim V = \text{tr deg}_F(F(V)), \text{ the transcendency degree of } F(V) \text{ over } F,$$

where $F(V)$ is the field of rational functions of V . More generally, the dimension of any nonempty affine subvariety $V \subseteq \mathbb{A}^n$ is defined as the maximum of the dimensions of its irreducible components. (For the sake of completeness, we can define the dimension of the empty variety to be -1 .) \square

We can give an identical definition of the dimension of an arbitrary subset $X \subseteq \mathbb{A}^n$. This can be handy even if our primary interest is in the dimension of a variety V , because the calculation of $\dim V$ can sometimes involve calculating dimensions of sets that are not varieties. Note that, by Proposition 7.4.7, $\dim X = \dim \bar{X}$ where (as usual) the overline denotes Zariski closure.

Example 7.8.2

(1) Suppose F is an infinite field. Then $I(\mathbb{A}^n) = (0)$ and therefore the co-ordinate ring of \mathbb{A}^n is $F[\mathbb{A}^n] \cong F[x_1, x_2, \dots, x_n]/(0) \cong F[x_1, x_2, \dots, x_n]$. The latter has the field $F(x_1, x_2, \dots, x_n)$ of rational functions as its field of quotients, and clearly its transcendency degree over F is n (because $\{x_1, x_2, \dots, x_n\}$ is a transcendence base). Hence, $\dim \mathbb{A}^n = n$.

On the other hand, over any field F , a finite subset $V = \{v_1, v_2, \dots, v_k\}$ of \mathbb{A}^n has dimension 0.¹⁸ In fact, since the $\{v_i\}$ are the irreducible components of V , it suffices to show points have dimension 0. However, for $v = (a_1, a_2, \dots, a_n) \in \mathbb{A}^n$, by Theorem 7.7.2 we have $I(\{v\}) = (x_1 - a_1, x_2 - a_2, \dots, x_n - a_n)$, the ideal of $F(x_1, x_2, \dots, x_n)$ generated by $x_1 - a_1, x_2 - a_2, \dots, x_n - a_n$. Consequently the co-ordinate ring $F[\{v\}] \cong F[x_1, x_2, \dots, x_n]/(x_1 - a_1, x_2 - a_2, \dots, x_n - a_n) \cong F$. Hence, $\dim\{v\} = \text{tr}_F \deg(F) = 0$.

Conversely, if a subset V of \mathbb{A}^n has dimension 0, then V must be finite. This will follow quickly from our next proposition because if V is infinite, then one of the co-ordinate functions x_1, x_2, \dots, x_n will be algebraically independent on V , implying $\dim V \geq 1$.

(2) Not surprisingly, isomorphic varieties must have the same dimension because their co-ordinate rings are isomorphic (see Remark 7.7.4 (3)).

18. When F is finite, algebraic geometry dimension of (nonempty) subvarieties V of \mathbb{A}^n is therefore trivial: $\dim V = 0$ for all V .

(3) Consider the two irreducible affine varieties in Example 7.7.5 (where F is algebraically closed):

$$V = \{(x, y) \in \mathbb{A}^2 : y = x^2\},$$

$$W = \{(x, y) \in \mathbb{A}^2 : y^2 = x^3\}.$$

The co-ordinate ring of the first is $F[V] \cong F[x]$, so $\dim V = \text{tr}_F \deg(F(x)) = 1$. The co-ordinate ring of W is $F[W] \cong F[x, y]/(y^2 - x^3)$. If we let \bar{f} denote the image of a polynomial $f(x, y)$ in the factor ring $F[x, y]/(y^2 - x^3)$, we see that \bar{x} is transcendental over F and \bar{y} is algebraic over $F[\bar{x}]$ (because $\bar{y}^2 = \bar{x}^3$). It follows that $\{\bar{x}\}$ is a transcendence base for the corresponding field of fractions. Thus, $\dim W = 1$ also. Here is an instance of where two varieties V and W have the same dimension (even 1) but are not isomorphic, in sharp contrast with vector space dimension (where vector spaces of the same dimension must be isomorphic as vector spaces).

(4) If V is a vector subspace of F^n , where F is an infinite field, then its algebraic geometry dimension and vector space dimension agree. To this end, suppose that V has vector space dimension m . Then there is a vector space isomorphism $T : F^n \rightarrow F^n$ such that $T(V) = F^m \subseteq F^n$. Since T and T^{-1} are given by right multiplication by associated $n \times n$ matrices on $1 \times n$ row vectors, they are inverse polynomial maps of \mathbb{A}^n . Their restrictions to V and F^m are therefore inverse polynomial maps, whence $V \cong \mathbb{A}^m$ as varieties. Since, by (1), the (algebraic geometry) dimension of \mathbb{A}^m is m , the algebraic geometry dimension of V is also m by (2).

Moreover, nothing is lost in translation: for a translate $W = \mathbf{b} + V$ of V , where $\mathbf{b} = (b_1, \dots, b_n) \in F^n$, we have the inverse polynomial maps $V \rightarrow W$, $\mathbf{v} \mapsto \mathbf{b} + \mathbf{v}$, and $W \rightarrow V$, $\mathbf{w} \mapsto -\mathbf{b} + \mathbf{w}$. Thus, V and W are isomorphic varieties and, therefore, $\dim W = \dim V = m$.

(5) Consider the real variety $V = V((y + x - 2)(y - 2x + 1)) \subseteq \mathbb{A}^2$, whose graph is drawn in Figure 7.5.

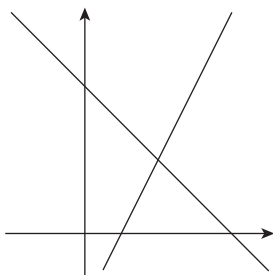


Figure 7.5 $V((y + x - 2)(y - 2x + 1))$

What is its dimension? At first blush, it might appear that V has dimension 2 because the two lines $y = -x + 2$ and $y = 2x - 1$ making up V are “independent.” But that is wrong. The two lines are the irreducible components of V , both of dimension 1 (they are translates of a 1-dimensional vector subspace), whence $\dim V = 1$. \square

Every spanning set S of a vector space contains a vector space basis—one takes a maximal linearly independent subset of S . An analogous result for a field extension $E \supseteq K$ is that any subset S of E that generates E over K (that is, $E = K(S)$), contains a transcendence base B for E over K . One takes a maximal algebraically independent (over K) subset B of S . (Observe that the set $A = \{a \in E : a \text{ is algebraic over } K(B)\}$ is a subfield of E containing K and S , whence $A = E$.) We can apply this to the field of rational functions $F(V)$ of an irreducible subvariety V of \mathbb{A}^n . The field $F(V)$ is generated over F by the **co-ordinate functions** x_1, x_2, \dots, x_n of V , that is, the polynomial maps of V induced by the polynomials x_i . There is little danger of confusion¹⁹ here in using the same notation x_i for two different things (as an indeterminate, and as a polynomial map of V) because the context will make the proper meaning clear. But to be precise, the i th co-ordinate function x_i is the polynomial map $p_i : V \rightarrow \mathbb{A}^1$, $(a_1, a_2, \dots, a_n) \mapsto a_i$ that projects a point in V to its i th component. Now we can see that a maximal algebraically independent set B of co-ordinate functions is a transcendence base for $F(V)$. By definition, the number of elements of B is the dimension of V . We record this in the following proposition.

Proposition 7.8.3

Let V be an irreducible affine subvariety of \mathbb{A}^n . Then $\dim V$ is the largest number m of the co-ordinate functions x_1, x_2, \dots, x_n that are algebraically independent on V . In other words, $\dim V$ is the largest m such that, for some distinct $x_{i_1}, x_{i_2}, \dots, x_{i_m} \in \{x_1, x_2, \dots, x_n\}$, the only polynomial $f \in F[x_{i_1}, x_{i_2}, \dots, x_{i_m}] \subseteq F[x_1, x_2, \dots, x_n]$ for which

$$f(v) = 0 \text{ for all } v \in V$$

is $f = 0$.

The proposition allows our arguments involving dimension to be a little less stiff and formal but nevertheless quite rigorous. For instance, we can now see at a glance that the variety W in Example 7.8.2 (3) has dimension 1, because the co-ordinate function x is algebraically independent on W (no nonzero polynomials in x vanish on W) but the co-ordinate functions x, y are clearly algebraically

19. We have avoided the wording “When there is danger of no confusion,” as on p. 1 of Mitchell’s “*Theory of Categories*”!

dependent because $x^3 - y^2$ vanishes on W . The reader may even wish to take Proposition 7.8.3 as the definition of the dimension of an irreducible variety.

Example 7.8.4

(1) Let F be an algebraically closed field and let $f \in F[x_1, x_2, \dots, x_n]$ be a nonconstant polynomial. Then the hypersurface $V = V(f) \subseteq \mathbb{A}^n$ has dimension $n - 1$. To check this, it suffices to consider the case where f is irreducible, because by Corollary 7.7.3 (4) the irreducible components of $V(f)$ are the varieties determined by the irreducible factors of f . Certainly $\dim V(f) \leq n - 1$ because f expresses an algebraic dependency relation on the co-ordinate functions x_1, x_2, \dots, x_n on V . Since f is nonconstant, some x_i appears in f and without loss of generality we take this to be x_n . We will show x_1, x_2, \dots, x_{n-1} are algebraically independent on V , whence by Proposition 7.8.3 we must have $\dim V = n - 1$. Write f as a polynomial

$$f = f_0(x_1, x_2, \dots, x_{n-1}) + f_1(x_1, x_2, \dots, x_{n-1})x_n + \dots + f_k(x_1, x_2, \dots, x_{n-1})x_n^k$$

in x_n with coefficients from $F[x_1, x_2, \dots, x_{n-1}]$ and with f_k nonzero. Now let g be a nonzero polynomial in $F[x_1, x_2, \dots, x_{n-1}]$. Since F is infinite, we can choose $a_1, a_2, \dots, a_{n-1} \in F$ such that both $g(a_1, a_2, \dots, a_{n-1})$ and $f_k(a_1, a_2, \dots, a_{n-1})$ are nonzero. (Apply Lemma 7.2.5 to the nonzero polynomial $f_k g$.) Since F is algebraically closed, the polynomial

$$h(x) = f_0(a_1, a_2, \dots, a_{n-1}) + f_1(a_1, a_2, \dots, a_{n-1})x + \dots + f_k(a_1, a_2, \dots, a_{n-1})x^k$$

has a zero in F , say a_n . Now $f(a_1, a_2, \dots, a_n) = h(a_n) = 0$, so $(a_1, a_2, \dots, a_n) \in V$. But viewing g as a polynomial in x_1, x_2, \dots, x_n we know that g does not vanish at the point $(a_1, a_2, \dots, a_n) \in V$. Thus, g does not vanish on V . Hence, x_1, x_2, \dots, x_{n-1} are algebraically independent on V , which completes our argument.

(2) As a particular instance of (1), we can conclude that $\dim SL_n(F) = n^2 - 1$ because $SL_n(F)$ is the hypersurface of $M_n(F) \cong \mathbb{A}^{n^2}$ determined by the polynomial $\det(x_{ij}) - 1$. \square

Remark 7.8.5

The algebraically closed assumption is needed in part (1) of this example. For instance, over the reals \mathbb{R} , the hypersurface $V(f)$ in \mathbb{A}^3 determined by the polynomial $f = x^2 + y^2 + z^2$ is just the point $(0, 0, 0)$, which has dimension 0, not $3 - 1 = 2$. \square

The following result generalizes Example 7.8.4. Its proof requires a careful argument, which we could handle but it would take us a little far afield. We refer the interested reader to pp. 56–57 of Shafarevich's text *Basic Algebraic*

Geometry 1 or Theorem 9.2 of Milne's *Algebraic Geometry*. The theorem is the analogue of the linear algebra result that the dimension of the solution space of a homogeneous system of linear equations drops by at most 1 with the introduction of a new equation.

Theorem 7.8.6

Let F be algebraically closed and $X \subseteq \mathbb{A}^n$ be an irreducible subset. If a polynomial $f \in F[x_1, x_2, \dots, x_n]$ has a zero in X but does not vanish identically on X , then

$$\dim(X \cap V(f)) = \dim X - 1.$$

In fact, all the irreducible components of $X \cap V(f)$ have this dimension. Of course, if f does vanish on X , then $\dim(X \cap V(f)) = \dim X$. So in general we can say that $\dim(X \cap V(f)) \geq \dim X - 1$ iff f has a zero in X .

Corollary 7.8.7

Let X be an irreducible subset of \mathbb{A}^n over an algebraically closed field F , and let $V = V(f_1, f_2, \dots, f_r)$ be the subvariety of \mathbb{A}^n determined by r polynomials f_i from $F[x_1, x_2, \dots, x_n]$. If the f_i have a common zero in X , then

$$\dim(X \cap V) \geq \dim X - r.$$

(In fact, a stronger statement is true: all the irreducible components of $X \cap V$ have dimension at least $\dim X - r$.)

Proof

There is no loss of generality in assuming that none of the f_i vanishes identically on X , because we could simply omit those polynomials from our list but keep the same intersection. Let X_1, X_2, \dots, X_k be the irreducible components of $X \cap V(f_1)$. By Theorem 7.8.6, $\dim X_i = \dim X - 1$ for all i . Since

$$X \cap V = (X_1 \cap V(f_2, f_3, \dots, f_r)) \cup \dots \cup (X_k \cap V(f_2, f_3, \dots, f_r))$$

expresses $X \cap V$ as a union of closed subsets, it follows that $\dim(X \cap V) = \dim(X_i \cap V(f_2, f_3, \dots, f_r))$ for some i . Clearly, $X_i \cap V(f_2, f_3, \dots, f_r)$ is nonempty because $X \cap V$ is nonempty. By induction on r , we can assume that $\dim(X_i \cap V(f_2, f_3, \dots, f_r)) \geq \dim X_i - (r - 1)$. Hence,

$$\begin{aligned} \dim(X \cap V) &= \dim(X_i \cap V(f_2, f_3, \dots, f_r)) \\ &\geq \dim X_i - (r - 1) \\ &= (\dim X - 1) - (r - 1) \\ &= \dim X - r, \end{aligned}$$

which establishes the corollary. For a proof of the stronger parenthetical statement, the reader can consult Corollary 9.7 of Milne's *Algebraic Geometry*. \square

Example 7.8.8

To illustrate Theorem 7.8.6, let us calculate the dimension of the variety N of all 2×2 nilpotent matrices over an algebraically closed field F . Consider the determinant and trace polynomials

$$\det(x_{11}, x_{12}, x_{21}, x_{22}) = x_{11}x_{22} - x_{21}x_{12}$$

$$\operatorname{tr}(x_{11}, x_{12}, x_{21}, x_{22}) = x_{11} + x_{22}.$$

The hypersurface $X = V(\det)$ of $M_2(F)$ determined by the polynomial \det is the variety of 2×2 singular matrices, which, as we noted in Example 7.4.20 is irreducible. By Theorem 7.8.6 (or Example 7.8.4 (1)), we see that $\dim X = 2^2 - 1 = 3$. Now a 2×2 matrix A is nilpotent exactly when $\det A = 0 = \operatorname{tr} A$ (because in terms of its characteristic polynomial we need $x^2 - (\operatorname{tr} A)x + \det A = x^2$). Hence, $N = X \cap V(\operatorname{tr})$. Since the trace function has a zero in X but doesn't vanish on X , by Theorem 7.8.6 we must have $\dim N = \dim X - 1 = 2$.

More generally, we show later in Lemma 7.10.3 that the variety of $n \times n$ nilpotents has dimension $n^2 - n$. \square

Proposition 7.8.9

- (1) For any subsets $X \subseteq Y$ of \mathbb{A}^n , $\dim X \leq \dim Y$.
- (2) If V is a subvariety of an irreducible variety W and $\dim V = \dim W$, then $V = W$.

Proof

(1) is clear from Proposition 7.8.3.²⁰ For (2), suppose $V \subseteq W$ are subvarieties of \mathbb{A}^n with W irreducible and $\dim V = \dim W = m$. By passing to an irreducible component of V of maximal dimension, we can assume that V is also irreducible. Without loss of generality, since $\dim V = m$, we can suppose that $\{x_1, x_2, \dots, x_m\}$ is a maximal algebraically independent set of co-ordinate functions of V . Since $V \subseteq W$ and $\dim V = \dim W$, $\{x_1, x_2, \dots, x_m\}$ must be a maximal algebraically independent set of co-ordinate functions of W as well. Inasmuch as $V \subseteq W$, we have $I(W) \subseteq I(V)$. Therefore, since both V and W are varieties, to show $V = W$ it suffices by Proposition 7.4.9 (4) to show that $I(V) \subseteq I(W)$. That is, if a polynomial $f \in F[x_1, x_2, \dots, x_n]$ vanishes on V , then f vanishes on W . On the contrary, suppose f does not vanish on W . Since $B = \{x_1, x_2, \dots, x_m\}$ is a transcendence base of $F(W)$ over F , the polynomial map of W induced by f is

20. Although Proposition 7.8.3 is stated for irreducible varieties, exactly the same proof works for irreducible subsets; alternatively we can take the Zariski closure, which does not alter the dimension.

algebraic over $F(B)$. Thus, there is a relationship

$$(*) a_0(x_1, x_2, \dots, x_m) + a_1(x_1, x_2, \dots, x_m)f + \dots + a_t(x_1, x_2, \dots, x_m)f^t = 0 \text{ on } W$$

in which the a_i are polynomials in $F[x_1, x_2, \dots, x_m]$ and a_0 is nonzero (start with the minimal polynomial of f over $F(B)$, which is irreducible). But a fortiori, $(*)$ holds on V , which implies $a_0(x_1, x_2, \dots, x_m)$ vanishes on V (because f does), contradicting the algebraic independence of x_1, x_2, \dots, x_m on V . Thus, f must vanish on W . \square

Remark 7.8.10

Part (2) of the proposition fails for irreducible subsets $X \subseteq Y$ with $\dim X = \dim Y$. For instance, $\dim GL_n(F) = \dim M_n(F) = n^2$ because $GL_n(F)$ is Zariski dense in $M_n(F) \cong \mathbb{A}^{n^2}$ (being a nonempty open subset of an irreducible variety). Also (2) fails for varieties $V \subseteq W$ if W is not irreducible. For instance, let V be a plane curve in \mathbb{R}^2 and let W be the union of V and another plane curve. Both V and W have dimension 1. \square

Proposition 7.8.11

If $X \subseteq \mathbb{A}^m$ and $Y \subseteq \mathbb{A}^n$, then

$$\dim(X \times Y) = \dim X + \dim Y.$$

Proof

We first establish the result when X and Y are irreducible subsets. By Proposition 7.4.16, $X \times Y$ is an irreducible subset of \mathbb{A}^{m+n} , so by Proposition 7.8.3 its dimension is the maximum number of co-ordinate functions that are algebraically independent on $X \times Y$. Let $\dim X = s$ and $\dim Y = t$. Then there is a set $S \subseteq \{x_1, x_2, \dots, x_m\}$ of s algebraically independent co-ordinate functions on X , and a set $T \subseteq \{x_1, x_2, \dots, x_n\}$ of t algebraically independent co-ordinate functions on Y . Let \overline{T} be the subset of $\{x_{m+1}, x_{m+2}, \dots, x_{m+n}\}$ corresponding to T upon replacing x_i by x_{m+i} . Then $S \cup \overline{T}$ is a set of $s + t$ algebraically independent co-ordinate functions on $X \times Y$. Hence, $\dim(X \times Y) \geq s + t$. Now let $U \subseteq \{x_1, x_2, \dots, x_{m+n}\}$ be an algebraically independent set of co-ordinate functions on $X \times Y$. The members of U that come from the first m co-ordinate functions must be independent on X and so are at most s in number. (If there was a dependence relation, it would hold on all of $X \times Y$ because members of Y are vacuously substituted in evaluating these co-ordinate functions.) Likewise, those members of U coming from the last n co-ordinate functions must be independent on Y and are at most t in number. Thus, U has at most $s + t$ members. This proves $\dim(X \times Y) = s + t$.

Now we treat the general case where X and Y are not necessarily irreducible. Let $X = X_1 \cup X_2 \cup \dots \cup X_k$ and $Y = Y_1 \cup Y_2 \cup \dots \cup Y_l$ be the decompositions of X and Y into their irreducible components. We can suppose X_1 and Y_1 have

maximal dimension among the respective components of X and Y . Thus, $\dim X = \dim X_1$ and $\dim Y = \dim Y_1$. Now

$$X \times Y = \bigcup_{i,j} (X_i \times Y_j)$$

decomposes $X \times Y$ as a union of closed subsets. Moreover, by Proposition 7.4.16, each $X_i \times Y_j$ is irreducible. Hence, the $X_i \times Y_j$ are the irreducible components of $X \times Y$. By the irreducible case of our present proposition, $\dim(X_i \times Y_j) = \dim X_i + \dim Y_j$. Therefore, $X_1 \times Y_1$ is a component of maximal dimension, whence

$$\begin{aligned} \dim(X \times Y) &= \dim(X_1 \times Y_1) \\ &= \dim X_1 + \dim Y_1 \\ &= \dim X + \dim Y. \end{aligned} \quad \square$$

The real power of vector space dimension arguments comes in their applications to a linear transformation $f : V \rightarrow W$. The most important result is “rank plus nullity equals $\dim V$,” where $\text{rank}(f) = \dim f(V)$ and $\text{nullity}(f) = \dim \ker(f)$. This invites the question: Is there an analogue of the result in our affine variety setting? The answer is “yes” and this is what will concern us for the remainder of the section. The linear transformations role is now taken over by the polynomial maps $f : V \rightarrow W$ from a subset V of affine n -space to a subset W of affine m -space. The image $f(V) = \{f(v) : v \in V\}$ is the same as for linear transformations, but we need an analogue of the kernel. For this, we consider the **fibre** $f^{-1}(w) = \{v \in V : f(v) = w\}$ of an element $w \in W$. In the case when V and W are varieties, fibres are subvarieties because they are inverse images of Zariski closed sets under a Zariski continuous function. Note, for a linear transformation $f : V \rightarrow W$, the kernel is the fibre $f^{-1}(0)$. But for the purposes of algebraic geometry, if f is a linear transformation of a subspace V of \mathbb{A}^n onto a subspace W of \mathbb{A}^m , then all fibres have the same algebraic geometry dimension because they are translates of $\ker(f)$ (see Example 7.8.2 (4)). Hence in this case, we can restate the rank, nullity connection in terms of algebraic geometry dimensions as:

$$\dim f^{-1}(w) = \dim V - \dim W \quad \text{for all } w \in W.$$

Here now is an extension of this to general affine spaces that will suit our purposes.

Theorem 7.8.12 (**Dimension of Fibres Theorem**)

Assume F is algebraically closed. Let $f : V \rightarrow W$ be a polynomial map of a subset $V \subseteq \mathbb{A}^n$ onto a subset $W \subseteq \mathbb{A}^m$. Then:

$$\dim f^{-1}(w) \geq \dim V - \dim W \text{ for some } w \in W.$$

Proof

Again we refer the interested reader to Shafarevich pp. 60–61 or Milne’s Chapter 10 for the full details, which would take us beyond our compass (although the basic ideas are within our reach). One first proves the result when V is irreducible. The idea is to reduce to a case for which some fibre $f^{-1}(w)$ is $V \cap V(f_1, f_2, \dots, f_r)$, where $r = \dim W$ and $f_1, f_2, \dots, f_r \in F[x_1, x_2, \dots, x_n]$ are polynomials with a common zero in $f^{-1}(w)$. By Corollary 7.8.7, $\dim f^{-1}(w) \geq \dim V - r = \dim V - \dim W$.

In the general case, let V_1 be an irreducible component of V of maximal dimension. Then $\dim V = \dim V_1$. By the irreducible case applied to the restriction $f : V_1 \rightarrow f(V_1)$ (and using Proposition 7.8.9 (1)) we have for some $w \in f(V_1)$

$$\begin{aligned} \dim f^{-1}(w) &\geq \dim V_1 - \dim f(V_1) \\ &\geq \dim V - \dim W. \end{aligned}$$

□

Remark. In the case that V and W are irreducible, the inequality in 7.8.12 actually holds for all $w \in W$, and equality holds on some nonempty open subset of W . Curiously, in the general case, the reverse inequality can hold for some fibre, but there may be no fibre for which the equality holds! □

We now have all the properties of algebraic geometry dimension that we require for our linear algebra applications in the next two sections.²¹ However, again for cultural reasons, we conclude this section with some remarks concerning other perspectives on dimension, which the reader may wish to check out in Hulek’s Chapter 3.

Remarks 7.8.13

Let V be an irreducible affine subvariety of \mathbb{A}^n over an algebraically closed field F . The dimension of V agrees with:

- (1) The **topological Krull dimension** of V relative to the induced Zariski topology. That is, $\dim V$ is the length l of the longest strictly descending chain

$$V = V_0 \supset V_1 \supset V_2 \supset \cdots \supset V_l$$

of nonempty irreducible closed subsets of V .

21. We now have the “Weyrwithal”!

- (2) The **ring-theoretic Krull dimension** of the co-ordinate ring $F[V]$ of V . That is, $\dim V$ is the length l of the longest strictly ascending chain

$$P_0 \subset P_1 \subset P_2 \subset \cdots \subset P_l$$

of prime ideals P_i of $F[V]$. (The equivalence of (1) and (2) follows from the inverse bijective maps V and I discussed in Section 7.7, so it depends ultimately on Hilbert's Nullstellensatz 7.3.3.)

- (3) The minimum of the vector space dimensions of the tangent spaces at points of V . □

7.9 GURALNICK'S THEOREM FOR $\mathcal{C}(3, n)$

We know that the variety $\mathcal{C}(k, n)$ of k -tuples of commuting $n \times n$ matrices over an algebraically closed field F is irreducible for $k = 1, 2$ and is irreducible for any k when $n < 4$, but is reducible for all $k \geq 4$ when $n \geq 4$. See Theorem 7.6.1, Corollary 7.6.7, and Proposition 7.6.8. The only open case is therefore $k = 3$, that is, the irreducibility of the variety $\mathcal{C}(3, n)$ of commuting triples of matrices. As of 2010, the current status of the problem is that $\mathcal{C}(3, n)$ is known to be irreducible for $n \leq 8$ over any algebraically closed field of characteristic zero and is known to be reducible (meaning not irreducible) for all $n \geq 29$ over any algebraically closed field. The question is still being actively researched, and it is likely that by the time this book goes to print, the gap $8 < n < 29$ will have been further closed, perhaps completely. The reducibility of $\mathcal{C}(3, n)$ for $n \geq 32$ was first established by Guralnick in 1992. In 2001, Holbrook and Omladič established reducibility for $n \geq 30$, using a nice refinement of Guralnick's proof. However, a minor change in their argument also establishes reducibility for $n = 29$. Since it was Guralnick who originally laid out the elegant algebraic geometry arguments, we shall informally still refer to the improved result as "Guralnick's theorem," and leave it to others to apportion due credit when the irreducibility question is finally settled, however that plays out.

In this section, we shall establish reducibility of $\mathcal{C}(3, n)$ for all $n \geq 29$ following the Holbrook and Omladič arguments, although we will use the Weyr form rather than the Jordan form to simplify centralizer arguments. We assume throughout that F is an algebraically closed field. It is interesting to note that Guralnick in his original argument did not use the Jordan form and instead started with a matrix A that was very close to a nilpotent Weyr matrix.²² Of course, our real interest in all of this is with the ASD connection that we established in Theorem 7.5.2. For Guralnick's theorem, when specialized to the

22. Guralnick's 4×4 block nilpotent matrix A , on p. 73 of his 1992 paper, can be put in Weyr form by conjugating with the permutation block matrix (23). Its Weyr structure is $(2s, s, s)$.

complex field \mathbb{C} , now shows that commuting triples of $n \times n$ complex matrices in general fail the ASD property for all $n \geq 29$. We are not aware of any proof of this that does not use algebraic geometry.

Lemma 7.9.1

If $\mathcal{C}(k, n)$ is irreducible, then its dimension is $n^2 + (k - 1)n$.

Proof

Let $V = \mathcal{C}(k, n)$. Let $Z = \{A \in M_n(F) : A \text{ is 1-regular}\}$. By Example 7.2.4, we know Z is a nonempty Zariski open subset of the irreducible variety $M_n(F)$, whence by Corollary 7.4.4, Z is an irreducible set with $\dim Z = \dim M_n(F) = n^2$. Let

$$M = \{(A_1, A_2, \dots, A_k) \in V : A_1 \in Z\}.$$

Observe, using 1-regularity, that for $(A_1, A_2, \dots, A_k) \in M$, we have each $A_i = p_i(A_1)$ for some uniquely determined polynomial $p_i \in F[x]$ of degree less than n (by Proposition 3.2.4). Thus, there is a natural affine isomorphism of $Z \times \mathbb{A}^n \times \mathbb{A}^n \times \dots \times \mathbb{A}^n$ ($k - 1$ copies of \mathbb{A}^n) onto M , namely,

$$(A_1, X_2, X_3, \dots, X_k) \mapsto (A_1, p_2(A_1), p_3(A_1), \dots, p_k(A_1)),$$

where $p_i(x)$ is the polynomial whose coefficients are taken from the n -tuple X_i . In particular, M has the same dimension as the domain. Therefore, by Proposition 7.8.11, we have

$$\dim M = \dim Z + (k - 1) \dim \mathbb{A}^n = n^2 + (k - 1)n.$$

Since Z is open, we see that M is a nonempty Zariski open subset of V . Hence, if V is irreducible, then M is dense in V and so $\dim V = \dim M = n^2 + (k - 1)n$. \square

Notation. Let a and b be fixed positive integers and let $n = a + 3b$. Let $W \in M_n(F)$ be the $n \times n$ nilpotent Weyr matrix with Weyr structure $(a + b, b, b)$:

$$W = \left[\begin{array}{cc|c|c} 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 \\ \hline & & 0 & I \\ & & & 0 \end{array} \right].$$

To assist the reader in keeping track of the size of the various matrix entries, we note that the four diagonal zeros of W are, respectively, of size $b \times b$, $a \times a$, $b \times b$, and $b \times b$.

Remark. Holbrook and Omladič used the Jordan form of W , which by Theorem 2.4.1 has the dual Jordan structure $(3, 3, \dots, 3, 1, 1, \dots, 1)$ with b lots of 3's and a lots of 1's. The centralizer of the Jordan matrix is not as easily handled as that of the Weyr matrix in the following argument. \square

By Proposition 3.2.2, we know that $\dim \mathcal{C}(W) = (a + b)^2 + b^2 + b^2 = a^2 + 2ab + 3b^2$. From our work in Chapter 2, we also know the precise form of the matrices that centralize W (see Proposition 2.3.3). We will write these as blocked matrices having the same block structure as W . Now let Γ be the variety consisting of all pairs (K, K') of matrices from $\mathcal{C}(W)$ of the form

$$K = \left[\begin{array}{cc|cc} 0 & A & B & C \\ 0 & 0 & 0 & D \\ \hline & & 0 & B \\ & & & 0 \end{array} \right], \quad K' = \left[\begin{array}{cc|cc} 0 & A' & B' & C' \\ 0 & 0 & 0 & D' \\ \hline & & 0 & B' \\ & & & 0 \end{array} \right]$$

where

$$(*) \quad AD' + BB' = A'D + B'B.$$

Here the sizes of the matrices A, B, C, D (and their prime counterparts) are, respectively, $b \times a$, $b \times b$, $b \times b$, $a \times b$. As the reader can easily check, the $(*)$ requirement is precisely the condition that K and K' commute. Without $(*)$, the set of all pairs (K, K') can be regarded as an irreducible subvariety of \mathbb{A}^{2n^2} of dimension $2(ba + b^2 + b^2 + ab) = 4ab + 4b^2$. (The variety is naturally affine isomorphic to $\mathbb{A}^{(4ab+4b^2)}$.) We can view $(*)$ as intersecting this subvariety with a collection of b^2 hypersurfaces coming from equating the entries of the $b \times b$ matrices of each side of the equation $(*)$. This intersection is nonempty because there are pairs (K, K') satisfying $(*)$. Remember that each nonempty intersection with a hypersurface drops the dimension by at most 1 (see Corollary 7.8.7). Thus,

$$\dim \Gamma \geq (4ab + 4b^2) - b^2 = 4ab + 3b^2.$$

It is not clear whether Γ is irreducible, but we don't need to know this.

Lemma 7.9.2

Let $f : \Gamma \times \mathbb{A}^3 \times SL_n(F) \longrightarrow \mathcal{C}(3, n)$ be the polynomial map defined by

$$\begin{aligned} & (K, K', \lambda_1, \lambda_2, \lambda_3, P) \\ \longmapsto & (P^{-1}(W + \lambda_1 I)P, P^{-1}(K + \lambda_2 I)P, P^{-1}(K' + \lambda_3 I)P). \end{aligned}$$

Let U be the image of f . Then $\dim U \geq n^2 + 2ab - a^2 + 3$.

Proof

The $(*)$ condition on pairs (K, K') in Γ ensures K and K' commute, while their general form ensures each of them commutes with W (Proposition 2.3.3). Therefore, (W, K, K') is a commuting triple of $n \times n$ matrices. Adding scalar matrices to W, K, K' won't alter this and neither will simultaneously conjugating the resulting matrices by some $P \in SL_n(F)$. Thus, f does indeed map into $\mathcal{C}(3, n)$. For $P \in SL_n(F)$, its inverse is the adjoint matrix $\text{adj } P$ whose entries are polynomial in the entries of P . Therefore, f is indeed a polynomial map.

We shall establish the inequality for the dimension of U with the aid of the Dimension of Fibres Theorem 7.8.12. Fix a member $T \in U$, say $T = f(X)$ for some $X = (K, K', \lambda_1, \lambda_2, \lambda_3, P)$. Now let $Y \in f^{-1}(T)$, say, $Y = (L, L', \mu_1, \mu_2, \mu_3, Q)$. Equating the first components of $f(X)$ and $f(Y)$, we have

$$P^{-1}(W + \lambda_1 I)P = Q^{-1}(W + \mu_1 I)Q.$$

Hence, by equating eigenvalues of each side, we have $\lambda_1 = \mu_1$. Therefore QP^{-1} centralizes $W + \lambda_1 I$. Thus, for the record, we have:

$$(i) \quad \lambda_1 = \mu_1 \text{ and } QP^{-1} \text{ centralizes } W.$$

Similarly, by equating the second and third components of $f(X)$ and $f(Y)$ (and equating eigenvalues, noting that K, K', L, L' are nilpotent), we have:

$$(ii) \quad \lambda_2 = \mu_2 \text{ and } (QP^{-1})K = L(QP^{-1}),$$

$$(iii) \quad \lambda_3 = \mu_3 \text{ and } (QP^{-1})K' = L'(QP^{-1}).$$

Conditions (i), (ii), and (iii) are also sufficient for $Y \in f^{-1}(T)$. Thus,

$$f^{-1}(T) = \{ (L, L', \lambda_1, \lambda_2, \lambda_3, Q) : \text{for some } Z \in \mathcal{C}(W) \cap SL_n(F), Q = ZP, \\ L = ZKZ^{-1}, L' = ZK'Z^{-1} \}.$$

Now remember that $K, K', \lambda_1, \lambda_2, \lambda_3, P$ are all fixed in this discussion. So the only freedom we have with the members of $f^{-1}(T)$ is in the choice of Q . But this is governed only by $Z \in C = \mathcal{C}(W) \cap SL_n(F)$. Hence, $\dim f^{-1}(T) = \dim C$. Formally, one checks this via the affine isomorphism

$$C \longrightarrow f^{-1}(T), \quad Z \longmapsto (ZKZ^{-1}, ZK'Z^{-1}, \lambda_1, \lambda_2, \lambda_3, ZP).$$

But from Theorem 7.8.6, since $SL_n(F)$ is a hypersurface of $M_n(F)$ which has a nonempty intersection with the irreducible variety $\mathcal{C}(W)$ but $\mathcal{C}(W) \not\subseteq SL_n(F)$, we

have $\dim C = \dim(\mathcal{C}(W) \cap SL_n(F)) = \dim \mathcal{C}(W) - 1 = a^2 + 2ab + 3b^2 - 1$.
Thus,

$$\dim f^{-1}(T) = a^2 + 2ab + 3b^2 - 1.$$

Let V be the domain of f . By Example 7.8.4 (2), $\dim SL_n(F) = n^2 - 1$. Hence, by Proposition 7.8.11,

$$\begin{aligned} \dim V &= \dim \Gamma + \dim \mathbb{A}^3 + \dim SL_n(F) \\ &\geq 4ab + 3b^2 + 3 + n^2 - 1 \\ &= 4ab + 3b^2 + n^2 + 2. \end{aligned}$$

Finally, via the Dimension of Fibres Theorem (7.8.12) we obtain our desired estimate for $\dim U$:

$$\begin{aligned} \dim U &\geq \dim V - \dim f^{-1}(T) \quad \text{for some } T \in U \\ &\geq 4ab + 3b^2 + n^2 + 2 - (a^2 + 2ab + 3b^2 - 1) \\ &= n^2 + 2ab - a^2 + 3. \end{aligned} \quad \square$$

Theorem 7.9.3 (Guralnick; Holbrook and Omladić)

The variety $\mathcal{C}(3, n)$ of commuting triples of $n \times n$ matrices over an algebraically closed field is reducible for all $n \geq 29$.

Proof

Let a and b be positive integers and let $n = a + 3b$. Assume $\mathcal{C}(3, n)$ is irreducible. Then by Lemma 7.9.1, $\dim \mathcal{C}(3, n) = n^2 + 2n$. Note that our subset U of commuting triples in Lemma 7.9.2 is contained in the proper (Zariski closed²³) subvariety X of $\mathcal{C}(3, n)$ consisting of commuting triples (A_1, A_2, A_3) in which A_1 has a repeated eigenvalue. (Having a repeated eigenvalue is a Zariski closed condition; see Example 7.1.11 (2).) Consequently, by Proposition 7.8.9, it follows that

$$\dim U \leq \dim X < \dim \mathcal{C}(3, n).$$

Hence, by Lemma 7.9.2 we must have

$$\begin{aligned} n^2 + 2n &> n^2 + 2ab - a^2 + 3 \\ \implies 6b + 2a &> 2ab - a^2 + 3 \\ \implies a^2 + 2(1 - b)a + (6b - 3) &> 0. \quad (**) \end{aligned}$$

23. For a reader who has dived directly into this section, it is important to note that subvarieties of \mathbb{A}^n are, according to our definition, Zariski closed subsets.

The quadratic $x^2 + 2(1 - b)x + (6b - 3)$ has real roots when $b \geq 8$, namely, $b - 1 - \sqrt{b^2 - 8b + 4}$ and $b - 1 + \sqrt{b^2 - 8b + 4}$. Therefore, every pair of positive integers a and b with $b \geq 8$ and

$$b - 1 - \sqrt{b^2 - 8b + 4} \leq a \leq b - 1 + \sqrt{b^2 - 8b + 4}$$

will contradict (**), and hence $\mathcal{C}(3, n)$ must be reducible for the corresponding value of $n = a + 3b$. A straightforward check reveals we can do this for all $n \geq 29$. For instance

$$\begin{array}{llll} a = 5, & b = 8 & \text{give} & n = 29; \\ a = 6, & b = 8 & \text{give} & n = 30; \\ a = 7, & b = 8 & \text{give} & n = 31; \\ a = 8, & b = 8 & \text{give} & n = 32. \end{array}$$

Observe how the interval determined by the roots of the quadratic grows as a function of b . In particular, the interval length is at least 4 for $b \geq 8$. Now make use of the observation that if $n = a + 3b$, then $n + 1 = (a + 1) + 3b$ and also $n + 1 = (a - 2) + 3(b + 1)$. For $b \geq 8$ and $n \geq 29$, one of these two forms must work for $n + 1$ if the form for n already works. Specifically, when $b \geq 8$ and

$$b - 1 - \sqrt{b^2 - 8b + 4} \leq a \leq b - 1 + \sqrt{b^2 - 8b + 4},$$

then either

$$\begin{aligned} b - 1 - \sqrt{b^2 - 8b + 4} \leq a + 1 \leq b - 1 + \sqrt{b^2 - 8b + 4}, \quad \text{or} \\ b - \sqrt{(b + 1)^2 - 8(b + 1) + 4} \leq a - 2 \leq b + \sqrt{(b + 1)^2 - 8(b + 1) + 4}. \end{aligned}$$

□

Corollary 7.9.4

For all $n \geq 29$, there exist three commuting $n \times n$ complex matrices that cannot be approximately simultaneously diagonalized.

Proof

This follows directly from Theorem 7.5.2 and Theorem 7.9.3. □

Notice that in deriving Corollary 7.9.4, we have used Theorem 7.5.2 only in the direction that the ASD property implies irreducibility of $\mathcal{C}(k, n)$. The Denseness Theorem 7.5.1 (whose proof we will give in Section 7.11) is not needed for this (it is used for the other direction). Over the complex field,

Theorem 7.5.2 still has the potential to establish ASD for commuting triples of low order $n \times n$ complex matrices. But to date, that hasn't happened. For instance the ASD property for commuting triples when $n = 5, 6, 7, 8$ has been established through the use of explicit perturbations. So currently we know the irreducibility of $\mathcal{C}(3, n)$ for $n = 5, 6, 7, 8$ only through this indirect route. See the discussion in Section 6.12 of Chapter 6.

7.10 COMMUTING TRIPLES OF NILPOTENT MATRICES

In the foregoing treatment of Guralnick's theorem, we used a nilpotent Weyr matrix W with Weyr structure $(a + b, b, b)$. In this section we allow W to have an arbitrary 3-tuple Weyr structure (n_1, n_2, n_3) . By reworking our earlier arguments, we construct sets of commuting triples of matrices that are of dimension even larger than previously. This and the use of other more general Weyr structures have the potential to further reduce the current²⁴ irreducibility–reducibility gap of $9 \leq n \leq 28$. The second part of this section briefly focuses on commuting triples of nilpotent matrices.

Some readers may choose to skip the somewhat more specialized material of this section. On the other hand, others may wish to see further testimony of our contention that the Weyr form is more suited to this type of analysis than the Jordan.

Here now is our more general approach.

Theorem 7.10.1

Let $n \geq 3$ be a fixed positive integer. A necessary condition for the variety $\mathcal{C}(3, n)$ of commuting triples of $n \times n$ matrices over an algebraically closed field to be irreducible is that the function

$$\Delta(n_1, n_2, n_3) = 2n + (n_1 - n_2)^2 + n_3(n - 3n_1) - 3$$

is positive for all partitions $n = n_1 + n_2 + n_3$ of n (where $n_1 \geq n_2 \geq n_3 \geq 1$).

Proof

Let W be the $n \times n$ nilpotent Weyr matrix with Weyr structure (n_1, n_2, n_3) .²⁵ (Note that our earlier structure $(a + b, b, b)$, where a, b are positive integers, is exactly

24. As of March, 2011.

25. This simple Weyr structure can have quite a long-winded Jordan counterpart, involving many 3×3 , 2×2 , and 1×1 basic Jordan blocks.

the case where $n_1 > n_2 = n_3$.) Relative to this block structure, let

$$K = \left[\begin{array}{cc|cc|c} 0 & A & B & C & D \\ & 0 & 0 & 0 & E \\ \hline & & & & B \\ \hline & & & & \end{array} \right], \quad K' = \left[\begin{array}{cc|cc|c} 0 & A' & B' & C' & D' \\ & 0 & 0 & 0 & E' \\ \hline & & & & B' \\ \hline & & & & \end{array} \right].$$

Here the diagonal blocks of K, K' have size $n_1 \times n_1$, $n_2 \times n_2$, and $n_3 \times n_3$. Blank entries are understood to be zeros, while the two displayed zeros in the top left block of K, K' are, respectively, $n_2 \times n_2$ and $(n_1 - n_2) \times (n_1 - n_2)$. Thus,

$$A, A' \text{ are } n_2 \times (n_1 - n_2),$$

$$B, B' \text{ are } n_2 \times n_3,$$

$$C, C' \text{ are } n_2 \times (n_2 - n_3),$$

$$D, D' \text{ are } n_2 \times n_3,$$

$$E, E' \text{ are } (n_1 - n_2) \times n_3.$$

Note that $K, K' \in \mathcal{C}(W)$ by Proposition 2.3.3, and the condition that K and K' commute is

$$(*) \quad AE' + [B \ C]B' = A'E + [B' \ C']B,$$

where both sides are $n_2 \times n_3$ matrices. Let

$$\Gamma = \{(K, K') : \text{subject to } (*)\}.$$

Then Γ is a variety and, by Corollary 7.8.7,

$$\begin{aligned} \dim \Gamma &\geq 2[n_2(n_1 - n_2) + n_2n_3 + n_2(n_2 - n_3) + n_2n_3 + (n_1 - n_2)n_3] - n_2n_3 \\ &= 2n_1n_2 + 2n_1n_3 - n_2n_3. \end{aligned}$$

Now, reworking the argument in Lemma 7.9.2, with the same definitions of V and U , we get

$$\begin{aligned}\dim V &= \dim \Gamma + \dim \mathbb{A}^3 + \dim SL_n(F) \\ &\geq (2n_1n_2 + 2n_1n_3 - n_2n_3) + 3 + (n^2 - 1) \\ &= 2n_1n_2 + 2n_1n_3 - n_2n_3 + n^2 + 2\end{aligned}$$

and

$$\begin{aligned}\dim U &\geq \dim V - \dim f^{-1}(T) \quad \text{for some } T \in U \\ &\geq 2n_1n_2 + 2n_1n_3 - n_2n_3 + n^2 + 2 - (n_1^2 + n_2^2 + n_3^2 - 1) \\ &= n^2 - (n_1 - n_2)^2 - n_3(n - 3n_1) + 3.\end{aligned}$$

(Note that we've used the formula in Proposition 3.2.2 that $\dim \mathcal{C}(W) = n_1^2 + n_2^2 + n_3^2$. Note also, when checking the description of members of $f^{-1}(T)$, that matrices of the form of K constitute an ideal of $\mathcal{C}(W)$, whence Γ is invariant under conjugation by any $Z \in \mathcal{C}(W) \cap SL_n(F)$.) Next, by the same argument as in Theorem 7.9.3, a necessary condition for $\mathcal{C}(3, n)$ to be irreducible is that $\dim U < \dim \mathcal{C}(3, n) = n^2 + 2n$. Hence, from our above estimate of $\dim U$, if $\mathcal{C}(3, n)$ is irreducible, then the function

$$\begin{aligned}\Delta(n_1, n_2, n_3) &= (n^2 + 2n) - [n^2 - (n_1 - n_2)^2 - n_3(n - 3n_1) + 3] \\ &= 2n + (n_1 - n_2)^2 + n_3(n - 3n_1) - 3\end{aligned}$$

must be positive. □

Note that in the special case of $n_1 = a + b$, $n_2 = b$, $n_3 = b$, our setup is the Holbrook and Omladič one and

$$\Delta(a+b, b, b) = 2(a+3b) + a^2 + b(-2a) - 3 = a^2 + 2(1-b)a + (6b-3).$$

Our earlier argument showing that $\mathcal{C}(3, 29)$ is reducible is a consequence of $\Delta(13, 8, 8) = 0$. We can also see this reducibility by noting that

$$\Delta(14, 8, 7) = 58 + 36 + 7(-13) - 3 = 0.$$

We note that for $n = 28$, the smallest value we can produce for $\Delta(a+b, b, b)$ is $\Delta(12, 8, 8) = 5$. On the other hand, in our revised construction (Theorem 7.10.1), we can reduce to $\Delta(13, 8, 7) = 56 + 25 + 7(-11) - 3 = 1$, almost zero. In fact, if we could show that the codimension of U in $\mathcal{C}(3, 28)$ is at least 2,

this would give reducibility for $n = 28$. (Here the **codimension** of a subset X in a set Y is simply $\dim Y - \dim X$.) This approach has the potential to establish reducibility for even smaller values of n .

Let $\mathcal{S}(3, n)$ denote the subset of $\mathcal{C}(3, n)$ consisting of triples of commuting $n \times n$ matrices (A_1, A_2, A_3) in which each A_i has a single eigenvalue (that is, A_i is a scalar matrix plus a nilpotent). Note that our constructed set U of commuting triples in Theorem 7.10.1 is a subset of $\mathcal{S}(3, n)$. Expanding on the above estimates of codimension, we see that the proof of the theorem establishes the following corollary:

Corollary 7.10.2

Fix $n \in \mathbb{N}$, $n \geq 3$. Suppose we have a lower bound estimate \mathcal{E}_n of the codimension of $\mathcal{S}(3, n)$ in $\mathcal{C}(3, n)$, that is,

$$\mathcal{E}_n \leq \dim \mathcal{C}(3, n) - \dim \mathcal{S}(3, n).$$

Then a necessary condition for $\mathcal{C}(3, n)$ to be irreducible is that

$$\Delta(n_1, n_2, n_3) = 2n + (n_1 - n_2)^2 + n_3(n - 3n_1) - 3 \geq \mathcal{E}_n$$

for all partitions $n = n_1 + n_2 + n_3$ of n (where $n_1 \geq n_2 \geq n_3 \geq 1$).

Note that, in terms of the estimates \mathcal{E}_n , our argument above for $n = 29$ takes $\mathcal{E}_{29} = 1$, and our proposed argument for $n = 28$ is that, maybe, we can take $\mathcal{E}_{28} = 2$.

In the remainder of this section, we briefly consider the corresponding question of irreducibility of the variety $\mathcal{CN}(3, n)$ of all triples of commuting $n \times n$ nilpotent matrices over an algebraically closed field. The irreducibility of this variety appears to be a somewhat stronger condition than irreducibility for $\mathcal{C}(3, n)$. As with Lemma 7.9.1, our first lemma is just as easily established more generally, namely for the variety $\mathcal{CN}(k, n)$ of all k -tuples of commuting $n \times n$ nilpotent matrices.

Lemma 7.10.3

If $\mathcal{CN}(k, n)$ is an irreducible variety, then its dimension is

$$n^2 + (k - 2)n - (k - 1).$$

Proof

Let N be the variety of all $n \times n$ nilpotent matrices. By Proposition 7.4.18, N is irreducible because

$$N = \bigcup_{T \in GL_n(F)} T^{-1}WT,$$

where W is the irreducible variety of all strictly upper triangular matrices (which is naturally isomorphic to $\mathbb{A}^{n(n-1)/2}$). The condition that $A \in M_n(F)$ be nilpotent can be given by n polynomial equations on its entries, namely, the vanishing of the n nonleading coefficients in its characteristic polynomial (its characteristic polynomial must be x^n). By Corollary 7.8.7, $\dim N \geq n^2 - n$. On the other hand, we have a strictly descending chain

$$N_1 = M_n(F) \supset N_2 \supset N_3 \supset \cdots \supset N_n \supset N,$$

where, for $j = 1, 2, \dots, n$,

$$N_j = \{A \in M_n(F) : A \text{ has an eigenvalue of multiplicity at least } j\}.$$

The N_j are irreducible by Proposition 7.4.18 because

$$N_j = \bigcup_{T \in GL_n(F)} T^{-1}W_jT,$$

where W_j is the irreducible variety of all upper triangular matrices

$$\begin{bmatrix} \lambda & * & * & \cdots & & * \\ 0 & \lambda & * & * & \cdots & * \\ 0 & & \ddots & & & \vdots \\ 0 & \cdots & 0 & \lambda & & * \\ 0 & & & 0 & * & \cdots & * \\ \vdots & & & & & \ddots & \vdots \\ 0 & \cdots & & & & 0 & * \end{bmatrix}$$

in which the first j diagonal entries are equal.

Claim: $\overline{N_j} \supset \overline{N_{j+1}}$ (strict containment) where the bar denotes Zariski closure.

For let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the eigenvalues of a matrix $A \in M_n(F)$, and let $p(x) \in F[x]$ be the characteristic polynomial. Let $p^{(i)}(x)$ be the i th formal derivative of $p(x)$. Fix an integer j with $1 \leq j \leq n$. The polynomial

$$s(\lambda_1, \lambda_2, \dots, \lambda_n) = p^{(j)}(\lambda_1)p^{(j)}(\lambda_2) \cdots p^{(j)}(\lambda_n)$$

is symmetric in the λ_i and therefore, by virtue of Proposition 7.1.10, we can express s as a polynomial $t(a_{11}, a_{12}, \dots, a_{nn})$ in the entries of A . Inasmuch as $p^{(j)}(\lambda_i) = 0$ for any eigenvalue λ_i of multiplicity at least $j + 1$, we see that t vanishes on N_{j+1} .

But clearly t doesn't vanish identically on N_j . Therefore, by Proposition 7.4.7, we cannot have N_{j+1} Zariski dense in N_j . Our claim is verified.

Since N_j is irreducible, so is \overline{N}_j by Proposition 7.4.3. We now have a strictly descending chain

$$N_1 = M_n(F) \supset \overline{N}_2 \supset \overline{N}_3 \supset \cdots \supset \overline{N}_n \supset N$$

of irreducible (closed) subvarieties of $M_n(F)$. Therefore, by Proposition 7.8.9 (2), we see that $\dim N \leq \dim M_n(F) - n = n^2 - n$. Thus, N is an irreducible variety of dimension $n^2 - n$.

Let R be the set of all 1-regular $n \times n$ nilpotent matrices (i.e., those with Weyr structure $(1, 1, \dots, 1)$). Then R is a nonempty Zariski open subset of N , whence R is dense in N because N is irreducible. Thus, $\dim R = \dim N = n^2 - n$. Also the nilpotent matrices that centralize a 1-regular nilpotent A are precisely the polynomials in A of degree up to $n - 1$ and with zero constant term. Since R is open, the set

$$X = \{(A_1, A_2, \dots, A_k) \in \mathcal{CN}(k, n) : A_1 \in R\}$$

is an open subset of $\mathcal{CN}(k, n)$. By exactly the same argument as in Lemma 7.9.1, if $\mathcal{CN}(k, n)$ is irreducible, then

$$\begin{aligned} \dim \mathcal{CN}(k, n) &= \dim X \\ &= \dim R + (k - 1)(n - 1) \\ &= (n^2 - n) + (k - 1)(n - 1) \\ &= n^2 + (k - 2)n - (k - 1). \end{aligned} \quad \square$$

In view of the following proposition, the estimates \mathcal{E}_n in Corollary 7.10.2 that we suggested for $n = 28, 29$ could well be quite conservative.

Proposition 7.10.4

If $\mathcal{C}(3, n)$ is irreducible, then $\mathcal{S}(3, n)$ has codimension at most $n - 1$ in $\mathcal{C}(3, n)$. However, the codimension will be exactly $n - 1$ if $\mathcal{CN}(3, n)$ is also irreducible.

Proof

The affine isomorphism

$$\begin{aligned} \mathbb{A}^3 \times \mathcal{CN}(3, n) &\longrightarrow \mathcal{S}(3, n) \\ (\lambda_1, \lambda_2, \lambda_3, A_1, A_2, A_3) &\longmapsto (\lambda_1 I + A_1, \lambda_2 I + A_2, \lambda_3 I + A_3) \end{aligned}$$

shows

$$\dim \mathcal{S}(3, n) = 3 + \dim \mathcal{CN}(3, n).$$

On the other hand, if we take

$$X = \{(A_1, A_2, \dots, A_k) \in \mathcal{CN}(k, n) : A_1 \in R\},$$

where R is the set of 1-regular nilpotent matrices, then using the argument in the proof of Lemma 7.10.3 for $k = 3$, we have

$$\dim \mathcal{CN}(3, n) \geq \dim X = n^2 + n - 2.$$

Therefore, by Lemma 7.9.1,

$$\dim \mathcal{C}(3, n) - \dim \mathcal{S}(3, n) \leq (n^2 + 2n) - (3 + n^2 + n - 2) = n - 1.$$

Thus, the codimension of $\mathcal{S}(3, n)$ in $\mathcal{C}(3, n)$ is at most $n - 1$. If $\mathcal{CN}(3, n)$ is also irreducible, then the codimension is exactly $n - 1$ by Lemma 7.10.3. \square

Klemen Šivic has informed us that he has established the following result using the Jordan form.²⁶ Here, we give a proof using (surprise, surprise) the Weyr form.

Theorem 7.10.5

The variety $\mathcal{CN}(3, n)$ of commuting triples of $n \times n$ nilpotent matrices is reducible for all $n \geq 13$. In particular, for each $n \geq 13$, there are commuting triples of $n \times n$ complex nilpotent matrices that cannot be perturbed to commuting nilpotent matrices, one of which is 1-regular.

Proof

We construct the set U of commuting triples as in the proof of Theorem 7.10.1 but this time taking all the λ_i to be zero. Let U_1 denote the resulting set of commuting triples of nilpotent matrices. Then

$$\dim U_1 \geq n^2 - (n_1 - n_2)^2 - n_3(n - 3n_1)$$

(3 less than the previous estimate for $\dim U$). Now assume that $n \geq 4$ and $\mathcal{CN}(3, n)$ is irreducible. Note that the first components of triples in U_1 all have rank at most $n - 2$, since they are similar to our fixed nilpotent Weyr matrix W with Weyr structure (n_1, n_2, n_3) , which has rank $n - n_1 \leq n - 2$. Thus, U_1 is contained in the proper subvariety of $\mathcal{CN}(3, n)$ consisting of the commuting nilpotent triples (A_1, A_2, A_3) where $\text{rank } A_1 \leq n - 2$. Consequently, by Lemma

26. Private communication in 2008.

7.10.3 and Proposition 7.8.9 (2), $\dim U_1 < \dim \mathcal{CN}(3, n) = n^2 + n - 2$. Thus, we see that a necessary condition for $\mathcal{CN}(3, n)$ to be irreducible is that the function

$$\begin{aligned}\Delta_1(n_1, n_2, n_3) &= (n^2 + n - 2) - [n^2 - (n_1 - n_2)^2 - n_3(n - 3n_1)] \\ &= n + (n_1 - n_2)^2 + n_3(n - 3n_1) - 2\end{aligned}$$

be strictly positive for all partitions $n = n_1 + n_2 + n_3$ of n (where $n_1 \geq n_2 \geq n_3 \geq 1$).

Specializing to the Holbrook and Omladić case, we have

$$\Delta_1(a + b, b, b) = a^2 + (1 - 2b)a + 3b - 2,$$

which is nonpositive for $b \geq 4$ and

$$(2b - 1 - \sqrt{4b^2 - 16b + 9})/2 \leq a \leq (2b - 1 + \sqrt{4b^2 - 16b + 9})/2.$$

Therefore, for such a and b the variety $\mathcal{CN}(3, n)$ is reducible for the corresponding $n = a + 3b$. One easily checks that this is so for all $n \geq 14$ (with $a = 2, b = 4$ giving the smallest case of $n = 14$).

Furthermore, for $n = 13 = 6 + 4 + 3$, we have

$$\Delta_1(6, 4, 3) = 13 + 2^2 + 3(13 - 3 \times 6) - 2 = 0,$$

so $\mathcal{CN}(3, 13)$ is also reducible.

We now establish the second statement of the theorem. Assume to the contrary that, for some $n \geq 13$, all commuting triples of $n \times n$ complex nilpotent matrices can be perturbed to commuting nilpotent matrices, one of which is 1-regular. Set

$$Y = \{(A_1, A_2, A_3) \in \mathcal{CN}(3, n) : A_1 \text{ is 1-regular}\}.$$

By our now standard argument, Y is an irreducible subset of $\mathcal{CN}(3, n)$. If we can show that Y is Euclidean dense in $\mathcal{CN}(3, n)$, then $\mathcal{CN}(3, n)$ will also be irreducible, establishing the anticipated contradiction. To this end, let $B = (B_1, B_2, B_3)$ be an arbitrary triple in $\mathcal{CN}(3, n)$. By assumption, B can be perturbed to a triple $(A_1, A_2, A_3) \in \mathcal{CN}(3, n)$, where one of the A_i is 1-regular, say, A_2 . From 1-regularity, A_1 is in this case a polynomial in $R = A_2$, say,

$$A_1 = a_1 R + a_2 R^2 + \cdots + a_{n-1} R^{n-1},$$

where each $a_i \in \mathbb{C}$. Since 1-regularity is an open condition, we can perturb these a_i to $\bar{a}_i \in \mathbb{C}$ such that

$$\bar{A}_1 = \bar{a}_1 R + \bar{a}_2 R^2 + \cdots + \bar{a}_{n-1} R^{n-1}$$

is 1-regular nilpotent. (As observed in the proof of Theorem 7.6.1, the 1-regularity of a matrix is equivalent to the nonvanishing of one of a finite number of polynomials p_1, \dots, p_k in the entries of the matrix. In our situation, we view R as being fixed and then the p_i are polynomials in a_1, \dots, a_{n-1} .) Notice that \bar{A}_1, A_2, A_3 still commute because they are all polynomials in $R = A_2$. Thus, we have perturbed B to a member of Y , which shows that Y is indeed Euclidean dense in $\mathcal{CN}(3, n)$. Our proof is complete. \square

A key observation used in Chapter 6, when studying the ASD question for commuting matrices, was that it is enough to handle commuting nilpotent matrices. Theorems 7.9.3 and 7.10.5 suggest that it is possible (perhaps even likely) that, for some n , the variety $\mathcal{C}(3, n)$ is irreducible while $\mathcal{CN}(3, n)$ is reducible. Over the complex field, this would mean that triples (A_1, A_2, A_3) of commuting $n \times n$ matrices can always be perturbed to commuting 1-regular matrices (e.g., diagonalizable with distinct eigenvalues), but, on the other hand, some triples (A_1, A_2, A_3) of commuting $n \times n$ nilpotent matrices can't be perturbed to commuting 1-regular nilpotent matrices. On the face of it, this appears to run counter to our “reduction to the nilpotent case” principle. However, the end goal of ASD is obtaining simultaneously diagonalizable matrices, which are never nilpotent unless zero.

7.11 PROOF OF THE DENSENESS THEOREM

We are grateful to S. Paul Smith of the University of Washington, Seattle, for supplying the proof of the Denseness Theorem 7.5.1 that we present in this section. To help the reader appreciate the “ring-craft” displayed in Paul Smith’s proof, we begin with two examples. The first is a simple proof that the Denseness Theorem holds in the case of the full affine space \mathbb{A}^n , even over the real field. The second example warns us that such a naïve approach will not work in general!

Example 7.11.1

Over the real or complex field F , every nonempty Zariski open subset U of affine n -space \mathbb{A}^n is Euclidean dense in \mathbb{A}^n . (Recall from Proposition 7.4.2 that \mathbb{A}^n is certainly an irreducible variety.) We establish this by induction on n . The result is trivial for $n = 1$ because U is then cofinite. Suppose $n > 1$. It is enough to establish

Euclidean denseness for the nonvanishing set

$$U = U(f) = \{(a_1, a_2, \dots, a_n) \in \mathbb{A}^n : f(a_1, a_2, \dots, a_n) \neq 0\}$$

of every nonconstant polynomial $f(x_1, \dots, x_n) \in F[x_1, \dots, x_n]$ (because by Proposition 7.2.1 these form the nonempty sets in a basis of open subsets for the Zariski topology). Without loss of generality, we can assume f has positive degree k in x_n so that

$$f(x_1, \dots, x_n) = f_0(x_1, \dots, x_{n-1}) + f_1(x_1, \dots, x_{n-1})x_n + \dots + f_k(x_1, \dots, x_{n-1})x_n^k,$$

where the $f_i \in F[x_1, \dots, x_{n-1}]$ and f_k is nonzero. Let $(b_1, b_2, \dots, b_n) \in \mathbb{A}^n$. We wish to ϵ -perturb (b_1, b_2, \dots, b_n) to a point in U . By induction, we may assume that the nonvanishing set $U(f_k)$ is Euclidean dense in \mathbb{A}^{n-1} . Therefore, we can perturb $(b_1, b_2, \dots, b_{n-1})$ to $(\bar{b}_1, \bar{b}_2, \dots, \bar{b}_{n-1})$ such that $f_k(\bar{b}_1, \bar{b}_2, \dots, \bar{b}_{n-1}) \neq 0$. Let

$$g(x_n) = f(\bar{b}_1, \bar{b}_2, \dots, \bar{b}_{n-1}, x_n) \in F[x_n].$$

Inasmuch as g is a nonzero polynomial in a single variable, it has only finitely many zeros. Hence, we can perturb b_n to \bar{b}_n such that $g(\bar{b}_n) \neq 0$. Now we have perturbed (b_1, b_2, \dots, b_n) to a member $(\bar{b}_1, \bar{b}_2, \dots, \bar{b}_n)$ of U , which establishes the Euclidean denseness of U . \square

Example 7.11.2

Despite the promise of Example 7.11.1, the Denseness Theorem fails in general over the real field \mathbb{R} . For let $V = V(f) \subseteq \mathbb{A}^2$ be the real variety determined by the irreducible polynomial

$$f(x, y) = y^2 - x^2(x - 1) \in \mathbb{R}[x, y].$$

Then V can be shown to be an irreducible real affine variety. However, unlike when the field is algebraically closed, and where we have the luxury of the Nullstellensatz 7.7.2, we can't just argue here that this is because f is an irreducible real polynomial—see Example 7.4.12. But for this particular polynomial f , one can show directly²⁷ that $I(V) = (f)$, whence $I(V)$ is a prime ideal (since f is irreducible)

27. We thank Keith Conrad for providing a proof of this. It is enough to show that if $g(x, y) \in \mathbb{R}[x, y]$ vanishes on V , then $f \mid g$. One writes $g(x, y) = f(x, y)q(x, y) + r(x, y)$ where $r(x, y) = u(x)y + v(x)$ for some $u(x), v(x) \in \mathbb{R}[x]$. Then one uses the fact that there are infinitely many first co-ordinates from points $(a, b) \in V$, and that $g(a, b) = 0$, to deduce that the polynomial $v(x)^2 - u(x)^2x^2(x - 1)$ must be zero because it vanishes at infinitely many $a \in \mathbb{R}$. Comparison of degrees in $v(x)^2 = u(x)^2x^2(x - 1)$ shows $u(x) = v(x) = 0$. Thus, $r(x, y) = 0$ and $f \mid g$.

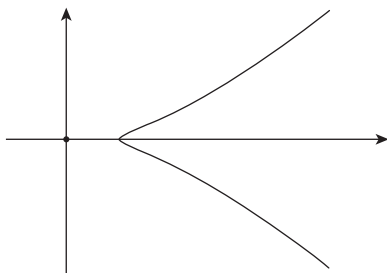


Figure 7.6 $f(x, y) = y^2 - x^2(x - 1)$

and therefore V is irreducible by Proposition 7.7.1 (2). Let $U = V \setminus \{(0, 0)\}$. Since points are Zariski closed, U is a nonempty Zariski open subset of V . But obviously U is not Euclidean dense in V because the graph of $y^2 = x^2(x - 1)$ in Figure 7.6 shows $(0, 0)$ is an isolated point: any open disc centered at $(0, 0)$ and with radius less than 1 is disjoint from U . \square

Now for the proof of:

The Denseness Theorem. *Nonempty Zariski open subsets of an irreducible complex affine variety V are Euclidean dense in V .*

Proof

Let U be a nonempty open subset of V and take Z to be its complement in V . It suffices to show that an arbitrary point $z \in Z$ is contained in the Euclidean closure of U .

By Noether's Normalization Theorem 7.3.6, there is an integer n and a quasi-finite surjective morphism $\pi : V \rightarrow \mathbb{C}^n$. Since π is a quasi-finite morphism, by Theorem 7.8.12 we have $\dim V = \dim \mathbb{C}^n = n$. By Proposition 7.8.9 (2) we have $\dim Z < \dim V$ because Z is proper (closed) subvariety of V . Thus, $\dim \pi(Z) \leq \dim Z < n$. In particular, the Zariski closure D of $\pi(Z)$ is a proper subset of \mathbb{C}^n . There is therefore a complex line L through $\pi(z)$ not contained in D . Observe that $\dim(D \cap L) < \dim L = 1$ by Proposition 7.8.9 (2), whence $D \cap L$ has dimension 0 and is therefore a finite set (see Example 7.8.2 (1)). Since π is a quasi-finite morphism, $\dim \pi^{-1}(L) = \dim L = 1$.

We have used the irreducibility of V when applying Noether's normalization. We again appeal to irreducibility of V to conclude that all the irreducible components of $\pi^{-1}(L)$ have dimension 1 (there are no singletons). To see this, view V as sitting inside some affine space \mathbb{C}^m so that π is a polynomial mapping of \mathbb{C}^m to \mathbb{C}^n . Note that since L is a translate of a 1-dimensional vector subspace of \mathbb{C}^n , there exist $n - 1$ linear polynomials $p_1, p_2, \dots, p_{n-1} \in \mathbb{C}[x_1, x_2, \dots, x_n]$ such that $L = V(p_1, p_2, \dots, p_{n-1})$. Let $f_i = p_i \circ \pi$ for $i = 1, \dots, n - 1$. These $f_i \in \mathbb{C}[x_1, x_2, \dots, x_m]$ and $\pi^{-1}(L) = V \cap V(f_1, f_2, \dots, f_{n-1})$. Hence, by Corollary 7.8.7, all the irreducible components of $\pi^{-1}(L)$ have dimension at least

$\dim V - (n - 1) = n - (n - 1) = 1$. Therefore, these irreducible components must have dimension exactly 1 because we know $\dim \pi^{-1}(L) = 1$.

Hence, the irreducible components of $\pi^{-1}(L)$ are curves, one of which, say C , passes through z . Now

$$\pi(C \cap Z) \subseteq \pi(C) \cap \pi(Z) \subseteq L \cap D,$$

so $\pi(C \cap Z)$ is finite. Inasmuch as π is a quasi-finite mapping, we see that $C \cap Z$ is a finite set.

However, every nonempty Euclidean open subspace of C is infinite: this is true when C is smooth because C is then a Riemann surface with respect to the Euclidean topology; for a C which is not necessarily smooth, there is a finite morphism $\alpha : C' \rightarrow C$ in which C' is a Riemann surface so, since α is continuous in the Euclidean topology, nonempty open subsets of C are infinite.

It follows from $C \cap Z$ being finite that the closure of $C \setminus Z$ in C relative to the Euclidean topology must be all of C , so contains z . In particular, z lies in the Euclidean closure of U , as desired. \square

Remark 7.11.3

It is certainly not true that the Zariski closure of an arbitrary subset U of an irreducible complex affine variety V agrees with its Euclidean closure. (For instance, \mathbb{Z} is Zariski dense in \mathbb{C} but, at last check, \mathbb{Z} is certainly not Euclidean dense.) However, what is true is that for any complex affine variety V , irreducible or otherwise, Zariski closure of any Zariski open subset U agrees with its Euclidean closure. We simply apply the Denseness Theorem to each of the nonempty $U \cap X_i$ where X_1, X_2, \dots, X_k are the irreducible components of V . \square

BIOGRAPHICAL NOTES ON HILBERT AND NOETHER

David Hilbert, a doyen of mathematics, was born on January 23, 1862, in Königsberg, now Kaliningrad, in Russia. He attended the university there and received his doctorate in 1885. His first work was on invariant theory and he proved his basis theorem for $R = F[x_1, x_2, \dots, x_n]$ in 1888. Using complicated computations, Paul Gordan had proved the theorem for two indeterminates 20 years earlier but his methods resisted generalization to more than two indeterminates. Hilbert's abstract existence approach to the problem was completely new: he proved that every ideal I of R has a finite generating set without actually constructing such a set. He submitted his result to *Mathematische Annalen* but Gordan refereed the paper, didn't appreciate the methods, and recommended its rejection. Hilbert, however, got wind of Gordan's appraisal and wrote to the editor, Felix Klein, defending his techniques. Klein accepted the paper without change, writing to Hilbert that

it was “*the most important work on general algebra that the Annalen has ever published.*” Hilbert was a staff member of the University of Königsberg from 1886 to 1895, becoming a professor in 1893. During that time he worked on algebraic number theory, resulting in a major report on work by Kummer, Kronecker, and Dedekind, but including a lot of Hilbert’s own concepts, which had a strong influence on the subject for many years. In 1895 he assumed the chair of mathematics at the University of Göttingen and his book *Grundlagen der Geometrie*, which appeared in 1899, is said to have had the greatest influence on geometry since Euclid. Hilbert’s address to the International Congress of Mathematicians in Paris in 1900 is undoubtedly one of the most influential speeches ever given on mathematics. The speech outlined 23 major mathematical problems for study in the twentieth century. They included the continuum hypothesis, the Riemann hypothesis, Goldbach’s conjecture, and the extension of Dirichlet’s principle. His later work on integral equations gave rise to research in functional analysis and the eponymous Hilbert space theory. He died on February 14, 1943, in Göttingen.

Emmy Noether was born on March 23, 1882, in Erlangen, Germany, as the daughter of the well-known algebraist Max Noether. In her late teens she planned to become a language teacher but, in 1900, instead decided to study mathematics at university. Being a woman, she had to obtain permission to sit in on courses from each of her professors and, with these given, she attended the University of Erlangen from 1900 to 1902. She then went to the University of Göttingen in 1903–1904, attending lectures by Hilbert, Klein, and Minkowski, before returning to Erlangen to work on her doctorate under Paul Gordan, which she completed in 1907. Although her research initially followed Gordan’s constructive methods, she soon came under the influence of Hilbert’s abstract approach (see the biographical sketch of Hilbert above). Hilbert and Klein invited her to return to Göttingen in 1915. However, again in part due to her gender, she was able to reach only the status of honorary professor there. In 1919 she began to focus on ideal theory. Her 1921 landmark publication *Idealtheorie in Ringbereichen* established the decomposition of ideals into intersections of primary ideals in a commutative ring with ascending chain condition on ideals, i.e., for what we now call commutative Noetherian rings. (This result is known as the Lasker–Noether Decomposition Theorem since it had been earlier established by Lasker for polynomial rings over a field.) Her normalization theorem appeared in 1926. However, with the Nazis coming to power, she was forced to leave Göttingen in 1933 and she spent her last two years at Bryn Mawr College in the United States, dying there on April 14, 1935. She actually published relatively little but was very generous with her ideas, particularly with her students. (Olga Taussky was one of her postdoctoral students at

Bryn Mawr.) To conclude, we quote from an article by Garrett Birkhoff: “If Emmy Noether could have been at the 1950 [International] Congress [of Mathematicians], she would have felt very proud. Her concept of algebra had become central in contemporary mathematics. And it has continued to inspire algebraists ever since.”

Bibliography

- Allman, E. S.; Rhodes, J. A., *Phylogenetic invariants for the general Markov model of sequence mutation*, Math. Biosci. **186** (2003), 113–144.
- , *Mathematical Models in Biology: An Introduction*, Cambridge University Press, Cambridge, 2004.
- Anderson, F. W.; Fuller, K. R., *Rings and Categories of Modules*, 2nd edition, Springer-Verlag, New York, 1992.
- Ara, P.; Goodearl, K. R.; O'Meara, K. C.; Pardo, E., *Separative cancellation for projective modules over exchange rings*, Israel J. Math. **105** (1998), 105–137.
- Baranovsky, V., *The variety of pairs of commuting nilpotent matrices is irreducible*, Transform. Groups **6** (2001), 3–8.
- Barría, J.; Halmos, P., *Vector bases for two commuting matrices*, Linear and Multilinear Algebra **27** (1990), 147–157.
- Basili, R.; Iarrobino, A., *Pairs of commuting nilpotent matrices, and Hilbert function*, J. Algebra **320** (2008), 1235–1254.
- Beidar, K. I.; O'Meara, K. C.; Raphael, R. M., *On uniform diagonalisation of matrices over regular rings and one-accessible regular algebras*, Comm. Algebra **32** (2004), 3543–3562.
- Belitskii, G., *Normal forms in matrix spaces*, Integral Equations Operator Theory **38** (2000), 251–283.
- Berman, A.; Plemmons, R. J., *A note on simultaneously diagonalizable matrices*, Math. Inequal. Appl. **1** (1998), 149–152.
- Bhatia, R.; Rosenthal, P., *How and why to solve the operator equation $AX - XB = Y$* , Bull. London Math. Soc. **29** (1997), 1–21.
- Birkhoff, G., (a) *The rise of modern algebra to 1936*, and (b) *The rise of modern algebra, 1936 to 1950*, in *Men and Institutions in American Mathematics*, ed. by D. Tarwater et al., Texas Tech. Press, 1976, pp. 41–63 and 65–85.
- Brandal, W.; *Commutative Rings Whose Finitely Generated Modules Decompose*, Lect. Notes Math. **723** Springer, Berlin, 1979.
- Brechenmacher, F., *Une histoire du théorème de Jordan de la décomposition matricielle*, doctoral thesis, Ecole des Hautes Etudes en Sciences Sociales (Paris), 2006.
- , *Les matrices : formes de représentations et pratiques opératoires (1850–1930)*, Site expert des Ecoles Normales Supérieures et du Ministère de l'Education Nationale (décembre 2006), <http://www.dma.ens.fr/culturemath/>. 65 pages.

- , *Une histoire de l'universalité des matrices mathématiques*, preprint, 2009.
- Brown, W. C., *Matrices over Commutative Rings*, Marcel Dekker, New York, 1993.
- , *Constructing maximal commutative subalgebras of matrix rings*, Rings, extensions, and cohomology (Evanston, IL, 1993), 35–40, Lect. Notes Pure Appl. Math., 159, Marcel Dekker, New York, 1994.
- , Call, F. W., *Maximal commutative subalgebras of $n \times n$ matrices*, Comm. Algebra **21** (1993), 4439–4460.
- Brualdi, R. A.; Pei, P.; Zhan, X., *An extremal sparsity property of the Jordan canonical form*, Linear Algebra Appl. **429** (2008), 2367–2372.
- Bryan, K.; Leise, T., *The \$25,000,000,000 eigenvector: The linear algebra behind Google*, SIAM Rev. **48** (2006), 569–581 (electronic).
- Camillo, V., *Row reduced matrices and annihilator semigroups*, Comm. Algebra **25** (1997), 176–1782.
- Cavender, J. A.; Felsenstein, J., *Invariants of phylogenies in a simple case with discrete states*, J. Class. **4** (1987), 57–71.
- Cayley, A., *A memoir on the theory of matrices*, Philos. Trans. R. Soc. Lond. **148** (1858), 17–37.
- Cherubino, S., *Sulle matrici permutabili con una data*, Rend. Sem. Mat. Padova **7** (1936), 128–156.
- Clark, J.; Brandal, W.; Barbut, E., *Decomposing finitely generated torsion modules*, Comm. Algebra **18** (1990), 225–245.
- Clark, J.; Lomp, C.; Vanaja, N.; Wisbauer, R., *Lifting modules. Supplements and projectivity in module theory*, Frontiers in Mathematics, Birkhäuser Verlag, Basel, 2006.
- Courter, R. C., *The dimension of maximal commutative subalgebras of K_n* , Duke Math. J. **32** (1965), 225–232.
- Cowsik, R. C., *A short note on the Schur-Jacobson theorem*, Proc. Amer. Math. Soc. **118** (1993), 675–676.
- Crilly, T., *Arthur Cayley. Mathematician Laureate of the Victorian Age*, Johns Hopkins University Press, Baltimore, MD, 2006.
- Dalecki, Ju. L., *On the asymptotic solution of a vector differential equation*, Dokl. Akad. Nauk SSSR **92** (1953), 881–884 (in Russian).
- de Boor, C.; Shekhtman, B., *On the pointwise limits of bivariate Lagrange projectors*, Linear Algebra Appl. **429** (2008), 311–325.
- Drazin, M. P.; Dungey, J. W.; Gruenberg, K. W., *Some theorems on commutative matrices*, J. London Math. Soc. **26** (1951), 221–228.
- Eisenbud, D., *Commutative Algebra with a View Toward Algebraic Geometry*, Springer-Verlag, New York, 1995.
- Ellerbroek, B. L.; Van Loan, C.; Pitsianis, N. P.; and Plemmons, R. J., *Optimizing closed-loop adaptive-optics performance with use of multiple control bandwidths*, J. Opt. Soc. Amer. A **11** (1994), 2871–2886.
- Erdos, J. A., *On products of idempotent matrices*, Glasgow Math. J. **8** (1967) 118–122.
- Felsenstein, J., *Inferring Phylogenies*, Sinauer Associates, Sunderland, MA, 2004.
- Fredholm, E. I., *Sur une classe d'équations fonctionnelles*, Acta Math. **27** (1903) 365–390.
- Frobenius, G., *Über lineare Substitutionen und bilineare Formen*, Crelle J. **84** (1878), 343–405.

- , *Über die vertauschbaren Matrizen*, S.-B. Deutsch. Akad. Wiss. Berlin Math.-Nat. Kl. (1896) 601–614.
- Gantmacher, F. R., *The Theory of Matrices. Vol. 1*, translated from the Russian by K. A. Hirsch; reprint of the 1959 translation, AMS Chelsea Publishing, Providence, 1998.
- Gerstenhaber, M., *On dominance and varieties of commuting matrices*, Ann. Math. **73** (1961), 324–348.
- Goodearl, K. R., *Von Neumann Regular Rings*, Pitman, London, 1979; 2nd edition, Krieger, Malabar, 1991.
- Guralnick, R. M., *A note on commuting pairs of matrices*, Linear and Multilinear Algebra **31** (1992), 71–75.
- ; Sethuraman, B. A., *Commuting pairs and triples of matrices and related varieties*, Linear Algebra Appl. **310** (2000), 139–148.
- Gustafson, W. H., *On maximal commutative algebras of linear transformations*, J. Algebra **42** (1976), 557–563.
- Hamilton, W. R., *Lectures on Quaternions*, Dublin, 1853.
- Han, Y., *Commuting triples of matrices*, Electron. J. Linear Algebra **13** (2005), 274–343 (electronic).
- Handbook of Linear Algebra*, edited by Leslie Hogben; associate editors: Richard Brualdi, Anne Greenbaum and Roy Mathias; Discrete Mathematics and its Applications (Boca Raton); Chapman & Hall/CRC, Boca Raton, 2007.
- Harima, T.; Watanabe, J., *The commutator algebra of a nilpotent matrix and an application to the theory of commutative Artinian algebras*, J. Algebra **319** (2008), 2545–2570.
- Hartley, B.; Hawkes, T. O., *Rings, Modules and Linear Algebra*, Chapman and Hall, London, 1970.
- Hershkowitz, D.; Schneider, H., *Height bases, level bases, and the equality of the height and the level characteristics of an M-matrix*, Linear and Multilinear Algebra **25** (1989), 149–171.
- , *Combinatorial bases, derived Jordan sets, and the equality of the height and the level characteristics of an M-matrix*, Linear and Multilinear Algebra **29** (1991a), 21–42.
- , *On the existence of matrices with prescribed height and level characteristics*, Israel J. Math. **75** (1991b), 105–117.
- Herzer, A.; Huppert, B., *Ein Satz von I. Schur über vertauschbare Matrizen*, Linear Algebra Appl. **71** (1985), 151–158.
- Hilbert, D., *Über die Theorie von algebraischen Formen*, Math. Ann. **36** (1890), 473–534.
- , *Über die vollen Invariantensysteme*, Math. Ann. **42** (1893), 313–373.
- Hoffman, K.; Kunze, R., *Linear Algebra*, 2nd edition, Prentice-Hall, Englewood Cliffs, 1971.
- Holbrook, J.; Omladič, M., *Approximating commuting operators*, Linear Algebra Appl. **327** (2001), 131–149.
- Horn, R. A.; Johnson, C. R., *Matrix Analysis*, Cambridge University Press, Cambridge, 1985; corrected reprint, 1990.
- ; *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

- Hulek, K., *Elementary Algebraic Geometry*, Student Mathematical Library, **20**; American Mathematical Society, Providence, RI, 2003.
- Ingraham, M. H.; Trimble, H. C., *On the matrix equation $TA = BT + C$* , Amer. J. Math. **63** (1941), 9–28.
- Jacobson, N., *Schur's theorems on commutative matrices*, Bull. Amer. Math. Soc. **50** (1944), 431–436.
- , *Basic Algebra. I*, W. H. Freeman and Company, San Francisco, 1974.
- , *Basic Algebra. II*, 2nd edition, W. H. Freeman and Company, San Francisco, 1980.
- Jordan, C., *Traité des Substitutions et des Équations Algébriques. Livre 2*, Gauthier-Villars, Paris, 1870.
- Kendig, K., *Elementary Algebraic Geometry*, Graduate Texts in Mathematics, **44**; Springer-Verlag, New York, 1977.
- Kleiner, I., *A History of Abstract Algebra*, Birkhäuser Boston, Boston, 2007.
- Košir, T.; Plestenjak, B., *On stability of invariant subspaces of commuting matrices*, Linear Algebra Appl. **342** (2002), 133–147.
- Laffey, T. J., *The minimal dimension of maximal commutative subalgebras of full matrix algebras*, Linear Algebra Appl. **71** (1985), 199–212.
- ; Lazarus, S., *Two-generated commutative matrix subalgebras*, Linear Algebra Appl. **147** (1991), 249–273.
- Lake, J. A., *A rate-independent technique for analysis of nucleic acid sequences: Evolutionary parsimony*, Mol. Biol. Evol. **4** (1987), 167–191.
- Lam, T. Y., *A theorem of Burnside on matrix rings*, Amer. Math. Monthly **105** (1998), 651–653.
- Lippert, R. A.; Strang G., *The Jordan forms of AB and BA* , Electronic J. of Linear Algebra **18** (2009), 281–288.
- Lumer, G.; Rosenblum, M., *Linear operator equations*, Proc. Amer. Math. Soc. **10** (1959) 32–41.
- McDuffee, C. C., *The Theory of Matrices*, Springer, Berlin, 1932.
- Milne, J. S., *Algebraic Geometry*, notes available electronically at <http://www.jmilne.org/math/CourseNotes/math631.html>.
- Moore, E. H., *On the reciprocal of the general algebraic matrix*, Bull. Amer. Math. Soc. **26** (1920), 394–395.
- , *General Analysis, Part I*, Mem. Amer. Philos. Soc., Philadelphia, 1935.
- Motzkin, T. S., *The Euclidean algorithm*, Bull. Amer. Math. Soc. **55** (1949), 1142–1146.
- ; Taussky, O., *Pairs of matrices with property L. II*, Trans. Amer. Math. Soc. **80** (1955), 387–401.
- Mumford, D., *Lectures on Curves on an Algebraic Surface*, Princeton University Press, Princeton, 1966.
- , *The Red Book of Varieties and Schemes*, Lect. Notes Math. **1358**, Springer-Verlag, New York, 1999.
- Murray, F. J.; von Neumann, J., *On rings of operators*, Ann. Math. **37**, (1936), 116–229.
- Neubauer, M. G.; Saltman, D., *Two-generated commutative subalgebras of $M_n(F)$* , J. Algebra **164** (1994), 545–562.
- ; Sethuraman, B. A., *Commuting pairs in the centralizers of 2-regular matrices*, J. Algebra **214** (1999), 174–181.

- Nicholson, W. K., *Introduction to Abstract Algebra*, 2nd edition, Wiley-Interscience, New York, 1999.
- , *Linear Algebra with Applications*, 6th edition, McGraw-Hill Ryerson, Whitby, Ontario, 2009.
- , *Semiregular modules and rings*, *Canad. J. Math.* **28** (1976), 1105–1120.
- Noether, E., *Idealtheorie in Ringbereichen*, *Math. Ann.* **83**, (1921), 24–66.
- , *Der Endlichkeitsatz der Invarianten endlicher linearer Gruppen der Charakteristik p* , *Nachr. Ges. Wiss. Göttingen* (1926), 28–35.
- Oblak, P., *The upper bound for the index of nilpotency for a matrix commuting with a given nilpotent matrix*, *Linear and Multilinear Algebra* **56** (2008), 701–711.
- O'Connor, J. J.; Robertson, E. F., articles in *MacTutor History of Mathematics*, available electronically at <http://www-history.mcs.st-andrews.ac.uk/Biographies>.
- O'Meara, K. C.; Vinsonhaler, C., *On approximately simultaneously diagonalizable matrices*, *Linear Algebra Appl.* **412** (2006), 39–74.
- O'Meara, K. C.; Vinsonhaler, C.; Wickless, W. J., *Identity-preserving embeddings of countable rings into 2-generator rings*, *Rocky Mountain J. Math.* **19** (1989), 1095–1105.
- Omladič, M., *A variety of commuting triples*, *Linear Algebra Appl.* **383** (2004), 233–245.
- Passman, D. S.; *A Course in Ring Theory*, Wadsworth & Brooks/Cole, Pacific Grove, 1991.
- Penrose, R. A., *A generalized inverse for matrices*, *Proc. Camb. Phil. Soc.* **51** (1955), 406–413.
- Phillips, H. B., *Functions of matrices*, *Amer. J. Math.* **41** (1919), 266–278.
- Reams, R., *Partitioned matrices*, in *Handbook of Linear Algebra*, 10-1–10-10.
- Reid, M., *Undergraduate Commutative Algebra*, Cambridge University Press, Cambridge, 1995.
- Ribenboim, P., *Fermat's Last Theorem for Amateurs*, Springer-Verlag, New York, 1999.
- Richman, D. J., *The singular graph of lower triangular, nilpotent matrices*, *Linear and Multilinear Algebra* **6** (1978/79), 37–49.
- ; Schneider, H., *On the singular graph and the Weyr characteristic of an M -matrix*, *Aequationes Math.* **17** (1978), 208–234.
- Roman, S., *Advanced Linear Algebra*, Graduate Texts in Mathematics, **135**; 3rd edition, Springer, Berlin, 2008.
- Rosenblum, M., *On the operator equation $BX - XA = Q$* , *Duke Math. J.* **23** (1956), 263–270.
- Schur, I., *Zur Theorie der vertauschbaren Matrizen*, *J. Reine Angew. Math.* **130** (1905), 66–76.
- Seemple, C.; Steel, M., *Phylogenetics*, Oxford Lecture Series in Mathematics and Its Applications, 24; Oxford University Press, Oxford, 2003.
- Sergeichuk, V. V., *Canonical matrices for linear matrix problems*, *Linear Algebra Appl.* **317** (2000), 53–102.
- Sethuraman B. A.; Šivic K., *Jet schemes of the commuting matrix pairs scheme*, *Proc. Amer. Math. Soc.* **137** (2009), 3953–3967.
- Shafarevich, I. R., *Basic Algebraic Geometry, Book 1, Varieties in Projective Space*, 2nd edition, Springer-Verlag, Berlin, 1994.

- Shapiro, H., *The Weyr characteristic*, Amer. Math. Monthly **106** (1999), 919–929.
- Shekhtman, B., *On a conjecture of Carl de Boer regarding the limits of Lagrange interpolants*, Constr. Approx. **24** (2006), 365–370.
- , *On a conjecture of Tomas Sauer regarding nested ideal interpolation*, Proc. Amer. Math. Soc. **137** (2008), 1723–1728.
- Šivic, K., *On varieties of commuting triples*, Linear Algebra Appl. **428** (2008), 2006–2029.
- Song, Y. K., *On the maximal, commutative, subalgebras of 14 by 14 matrices*, Comm. Algebra **25** (1997), 3823–3840.
- Spiegel, E., *Incidence algebras with involution*, Linear Algebra Appl. **405** (2005), 155–162.
- Strang, G., *Linear Algebra and Its Applications*, 4th edition, Thomson Brooks/Cole, 2006.
- Sturmfels, B.; Sullivant, S., *Toric ideals of phylogenetic invariants*, J. Computational Biology **12** (2005), 204–228.
- Suprunenko, D. A.; Tyshkevich, R. I., *Commutative Matrices*, Academic Press, New York, 1968.
- Sylvester, J., *Sur l'équation en matrices $px = xq$* , C. R. Acad. Sci. Paris **99** (1884) 67–71, 115–116.
- Turnbull, H. W.; Aitken, A. C., *An Introduction to the Theory of Canonical Matrices*, Blackie and Son Limited, London, 1952.
- Von Neumann, J., *On regular rings*, Proc. Nat. Acad. Sci. USA **23** (1937), 16–22.
- Wadsworth, A. R., *The algebra generated by two commuting matrices*, Linear and Multilinear Algebra **27** (1990), 159–162.
- Ware, R., *Endomorphism rings of projective modules*, Trans. Amer. Math. Soc. **155** (1971), 233–256.
- Warfield, R. B., Jr., *Exchange rings and decompositions of modules*, Math. Ann. **199** (1972), 31–36.
- Weintraub, S. H., *Jordan Canonical Form : Theory and Practice*, Morgan & Claypool, San Rafael, California, 2009.
- Weiss, E., *First Course in Algebra and Number Theory*, Academic Press, New York, 1971.
- Weyl, H., *The Classical Groups. Their Invariants and Representations*, 2nd edition, Princeton University Press, Princeton, 1946.
- Weyr, E., *Répartition des matrices en espèces et formation de toutes les espèces*, C. R. Acad. Sci. Paris **100** (1885), 966–969.
- , *Zur Theorie der bilinearen Formen*, Monatsh. Math. Physik **1** (1890), 163–236.
- Xue, W., *Characterization of rings using direct-projective modules and direct-injective modules*, J. Pure Appl. Algebra **87** (1993), 99–104.
- Yu, H.-P., *On strongly pi-regular rings of stable range one*, Bull. Austral. Math. Soc. **51** (1995), 433–437.

INDEX

- 1-generated subalgebra, 206
- 1-regular matrix, 8
- 2-correctable perturbation, 262, 268
- 2-generated subalgebra, 206
- 2-regular matrix, 276
- 3-generated commutative subalgebra, 226
- 3-regular matrix, 230, 300
- A^T , 64
- $V \cong W$, 314
- $\Delta(n_1, n_2, n_3)$, 370
- $\|A\|$, 249
- $\|\mathbf{a}\|$, 249
- \emptyset , 312
- $\mathbb{Z}/(n)$, 137
- \overline{X} , 329
- $\pi(n)$, 41
- \sqrt{J} , 350
- a^+ , 166
- a^\perp , 166
- f^* , 352
- \mathcal{E}_n , 373
- adjoint map, 159
- affine
 - n -space, 311
 - variety, 310, 311
- Aitken, A. C., 37, 81
- algebra
 - automorphism, 18
 - Banach, 153
 - C^* , 153
 - homomorphism, 16
 - over a commutative ring, 131
 - over a field, 8
 - quaternion, 132
 - von Neumann, 153, 197
- algebraic
 - element, 354
 - extension, 354
 - geometry dimension, 355
 - multiplicity, 6
- algebraically
 - closed field, 5
 - dependent, 354
 - independent, 354
- algorithm for Weyr form, 82
- Allman, E. S., 238, 239, 243, 246, 249
- \mathbb{A}^n , 311
- $\text{ann}(S)$, 170
- $\text{ann}(t)$, 170
- annihilator, 134
- annihilator ideal, 170, 229
- $\text{ann}_R(S)$, 134
- approximately simultaneously
 - diagonalizable, 239, **250**
- Ara, P., 153, 159
- ascending chain
 - condition, 326
 - of open subsets, 331
- ASD
 - matrices, 239, **250**
 - property, 310, 339
 - reduction principle, 259
- automorphism, 18

- Banach algebra, 153
- $B(a, r)$ open ball, 340
- Barbut, E., 173
- Barria, J., 210, 220
- basic
 - Jordan matrix, **38**, 49
 - Weyr matrix, 49, **50**
- basis
 - for a module, 144
 - orthonormal, 85
 - standard, 5
- $[\mathcal{B}', \mathcal{B}]$, 17
- Beidar, K. I., 160, 178, 189
- Belitskii algorithm, 80, 81
- Belitskii, G., 80
- Bergman, G. M., 194
- Berman, A., 238
- Bhatia, R., 33
- biographical note on
 - Cayley, A., 236
 - Frobenius, G., 123
 - Hamilton, W. R., 237
 - Hilbert, D., 381
 - Jordan, C., 42
 - Motzkin, T. S., 307
 - Noether, E., 382
 - Sylvester, J., 43
 - Taussky, O., 308
 - Von Neumann, J., 197
 - Weyr, E., 94
- Birkhoff, G., 383
- block, 12
 - diagonal, 15
 - lower triangular, 14
 - matrix, 12
 - upper triangular, 14
- blocked matrix, 12
- Box, G., 246
- Brandal, W., 173
- Brauer, R., 32
- Brechenmacher, F., 95
- Brown, W. C., 212, 267
- Brualdi, R., 40
- Burnside, W., 32, 108
- Burnside's orbit-counting theorem, 108
- Burnside's theorem on matrix subalgebras, 219
- $\mathcal{C}(A)$ for a matrix A , 65, 96
- $\mathcal{C}(\mathcal{A})$ for a subalgebra \mathcal{A} , 221
- $\mathcal{C}(k, n)$, 310, 315
- C^* -algebra, 153
- $\mathcal{C}(A_1, A_2, \dots, A_k)$, 265
- Call, F. W., 267
- Camillo, V., 36
- canonical form, 35
 - Jordan, 36, **39**
 - rational, 36
 - Weyr, 36, **61**
- Cavender, J. A., 246
- Cayley, A., 7, **236**
- Cayley–Hamilton
 - generalized equation, 211
 - theorem, 7, 236, 237
- Cayley's theorem, 129
- centralize, 112
- centralizer, 65, 96
 - of Jordan matrix, 97, 98
 - of subalgebra, 221
 - of Weyr matrix, **68**, 100
- chain condition
 - ascending, 326
 - descending, 332
- change of basis matrix, 17
- characteristic
 - polynomial, 6
 - Segre, 39
 - Weyr, 61
- Cherubino, S., 253
- Chinese remainder theorem, 210
- Clark, J., 173
- Clemens, H., 340
- closed set, 320
- closure, \overline{X} , 329
- $\mathcal{CN}(k, n)$, 373
- co-ordinate
 - function, 357
 - ring, 347
 - vector, 17
- codimension, 373

- column rank, 6
- commutative diagram, 134
- commutative subalgebra
 - 1-generated, 206
 - 2-generated, 216
 - 3-generated, 226
- commuting
 - pairs, 271
 - triples, 276
- companion matrix, 20, 26, 105
- complementary summand, 141
- completely reducible module, 149
- component, 136
 - irreducible, 333
- condition
 - ascending chain, 326
 - descending chain, 332
- conjugate partitions, 74
- conjugation, 18
- Conrad, K., 340, 379
- continuous map, 324
- Courter, R. C., 221, 267
- cyclic
 - submodule, 134
 - vector, 105

- Dalecki, Ju. L., 33
- de Boor, C., 239, 257
- decomposition
 - generalized eigenspace, 30
 - Goodearl, 184
 - Jordan, 184
 - primary, 29
 - singular value, 166
 - Weyr, 184
- dense subset, 329
- Denseness Theorem, 340, 369, 378
- descending chain
 - condition, 332
 - of closed subsets, 331
- $\det A$, determinant of matrix A , 6
- determinant map, 317
- diagonalizable
 - approximately simultaneously, 239, **250**
 - matrix, 22, 250
 - simultaneously, **250**
- diagram
 - commutative, 134
 - Ferrer's, 74
 - Young, 75
- Dieudonné, J., 197
- $\dim V$, vector space dimension, 5
- $\dim V$, algebraic geometry dimension, 355
- dimension
 - algebraic geometry, 355
 - Jordan centralizer, 99
 - ring-theoretic Krull, 364
 - topological Krull, 363
 - Weyr centralizer, **101**
- Dimension of Fibres Theorem, 363, 368
- direct sum
 - external, 135
 - internal, 136
 - of matrices, 15
 - of subspaces, 20
- direct summand, 141
- directly finite ring, 195
- divisible group, 128
- DNA sequence analysis, 241
- domain
 - principal ideal, 170
 - unique factorization, 172
- Doogue, G., xviii
- Drazin, M. P., 254
- dual partition, 74
- duality between canonical forms, **74**
- Duney, J. W., 254

- $E(\lambda)$, 6
- eagle, 307
 - kea*, 66
- eigenspace, 6
- eigenvalue, 6
- eigenvector, 6
- Eisenbud, D, 350
- element
 - algebraic, 354
 - irreducible, 172

- regular, 152
- torsion, 170
- transcendental, 354
- unit-regular, 195
- von Neumann regular, 152
- elementary row operations, 9
- $\text{End}_R(M)$, 133
- $\text{End}(M)$, 129
- endomorphism, 129, 133
- Erdos, J. A., 32
- Euclidean
 - metric, 249
 - topology, 321, 339, **340**
- exchange ring, 153
- extension
 - algebraic, 354
 - purely transcendental, 354
- external direct sum, 135

- $F[A_1, A_2, \dots, A_k]$, 8
- $F[A]$, 8
- $F[S]$, 8
- $F[x]$, 5
- F^n , 5
- factor module, 132
- Felsenstein, J., 243, 246
- Ferrer's diagram, 74
- FGC property, 173
- fibre, 325, 362
- field of rational functions, 354
- finite map, 328
- finitely generated submodule, 134
- form
 - Jordan, **39**
 - Weyr, **54**
- formula
 - Frobenius, 96, 99, 225, 262
 - Jordan centralizer dimension, 99
 - Lagrange interpolation, 254
 - leading edge dimension, 111
 - Weyr centralizer dimension, **101**
- Fredholm, E. I., 166
- free module, 144
- Frobenius formula, 96, 99, 225, 262
- Frobenius, G., 32, **123**, 209, 237
- full column-rank, 6
- full rank factorization, 167
- fundamental homomorphism
 - theorem, 133
- fundamental theorem for f.g. torsion
 - modules over a PID, 173
- $F[V]$, co-ordinate ring of V , 347
- $F(V)$, field of rational functions, 355

- g -block pullback, 227
- general linear group, 32, 58, 319
- generalized Cayley–Hamilton equation,
 - 211, 232
- generalized eigenspace, 28
 - decomposition, 30
- generalized inverse, 154
- generated by a_1, a_2, \dots, a_k , 134
- geometric multiplicity, 6
- Gerstenhaber's theorem, 122, **219**, 225,
 - 226, 234, 255, 273, **273**, 342, **343**
- Gerstenhaber, M., 96, 106, 219
- $GL_n(F)$, 32, 58, 125, 319
- Goodearl decomposition of
 - module, 184
- Goodearl, K. R., 125, 152, 153, 159, 178,
 - 194, 195, 340
- Google, 3
- Gordan, P., 381
- group
 - divisible, 128
 - general linear, 32, 58, 319
 - special linear, 317
 - symmetric, 18
 - torsion-free, 128
- Gruenberg, K. W., 254
- Guralnick's theorem, 364, **368**,
 - 369, 370
- Guralnick, R. M., 118, 119, 307, 309, 310,
 - 342, 364

- H-form, 44, 82, 239
- Halmos, P. R., 210, 220
- Hamilton, W. R., 7, **237**

- Han, Y., 240, 307
 Harima, T., 82
 Hartley, B., 173
 Hawkes, T. O., 173
 Hershkowitz, D., 62
 Hilbert space, 152, 153
 Hilbert's
 basis theorem, 310, 326, 331
 nullstellensatz, 310, 326, 327, 350, 364
 Hilbert, D., 310, 326, **381**
 Hoffman, K., 4, 105
 Holbrook, J., 307, 310, 364, 366, 368, 372, 377
 homeomorphism, 325
 homogeneous
 system, 6
 Weyr structure, 52, 277
 homomorphism
 R -, 133
 algebra, 16
 module, 133
 Horn, R. A., 4, 14, 80, 212
 Hulek, K., 312, 327, 350, 351, 363
 hypersurface, 312

 ideal
 annihilator, 229
 idempotently generated, 156
 maximal, 135
 minimal, 135, 138
 radical, 350
 radical of an, 350
 idempotent
 linear transformation, 27
 matrix, 24
 of a ring, 139
 idempotently generated
 ideal, 156
 indecomposable module, 142
 index of nilpotency, 7, 27
 induced topology, 328
 Ingraham, M. H., 210
 injection, 141
 internal direct sum, 136

 invariant subspace, 19
 inverse
 generalized, 154
 image, 180, 324
 Moore–Penrose, 152, 166–168
 quasi, 153
 irreducible
 component, 333
 element, 172
 module, 135
 subset, 328
 topological space, 328
 variety, 310
 isomorphic varieties, 314
 isomorphism
 module, 133
 of varieties, 314
 $I(W)$, 331

 Jacobson, N., 18, 124, 125, 173, 305, 318, 326, 354
 Johnson, C. R., 4, 14, 212
 Jordan
 basic matrix, **38**, 49
 canonical form, 36, **39**
 decomposition of module, 184
 form, 36, **39**
 centralizer, 98
 for a nilpotent endomorphism, 185
 matrix, **39**
 structure, **39**
 of general matrix, 39
 Jordan, C., **42**

 k -correctable perturbation, 261
 k -generated subalgebra, 203
 k -regular matrix, 8
 kea, 66
 kookaburra, 226
 ker T , 5
 kernel, 5
 Klein, F., 381
 kookaburra, 226
 eagle, 307

- Krull dimension
 - ring-theoretic, 364
 - topological, 363
- Kunze, R., 4, 105
- Laffey, T. J., 118, 210, 220, 221, 223, 224, 346
- Laffey–Lazarus theorem, 224, 225, 346
- Lagrange, J.L., 107
- Lagrange interpolation formula, 254
- Lagrange’s theorem, 107
- Lake, J. A., 246
- Lam, T. Y., 219
- Lang, S., 320
- Lazarus, S., 118, 210, 220, 221, 223, 224, 346
- leading edge
 - dimension formula, 111
 - subspace, 109
- left R -module, 127
- linear group
 - general, 319
 - special, 317
- Lippert, R. A., 62
- Lumer, G., 33
- $\mathcal{L}(V)$, 18
- map
 - adjoint, 159
 - continuous, 324
 - determinant, 317
 - finite, 328
 - homeomorphism, 325
 - polynomial, 310, 314
 - quasi-finite, 328
 - rational, 320
 - trace, 18, 317
- Maschke’s theorem, 125
- matrices
 - approximately simultaneously diagonalizable, 239, **250**
 - ASD, 239, **250**
 - commuting
 - pairs, 271
 - triples, 276
 - direct sum of, 15
 - similar, 18
 - unitarily, 85
 - simultaneously
 - diagonalizable, **250**
 - triangularized, 70, **71**
- matrix
 - 1-regular, 8
 - 2-regular, 276
 - 3-regular, 230
 - basic Jordan, **38**, 49
 - basic Weyr, 49
 - block, 12
 - block diagonal, 15
 - block lower triangular, 14
 - block upper triangular, 14
 - blocked, 12
 - canonical form, 35
 - change of basis, 17
 - companion, 20, 26, 105
 - diagonalizable, 22, 250
 - direct sum, 15
 - idempotent, 24
 - Jordan, **39**
 - k -regular, 8
 - minimal polynomial of, 25
 - nilpotent, 7, 27
 - nonderogatory, 8, 103, 105
 - norm, 249
 - of a transformation, 17
 - of an endomorphism, 145
 - partition, 12
 - permutation, 18, 48
 - quasi-inverse, 268
 - reduced row-echelon, 35, 50
 - strictly block upper triangular, 14
 - Toeplitz, 68
 - unitary, 85
 - Weyr, **54**
- maximal
 - commutative subalgebra, 112, 221
 - left ideal, 135
- McCoy, N. H., 253
- Milne, J. S., 37, 326, 359, 360, 363

- minimal
 - left ideal, 135, 138
 - polynomial, 25, 354
- Mitchell, B., 357
- $M_{m \times n}(R)$, 128
- $M_n(F)$, 6
- module
 - basis, 144
 - completely reducible, 149
 - endomorphism, 133
 - factor, 132
 - free, 144
 - homomorphism, 133
 - indecomposable, 142
 - irreducible, 135
 - isomorphism, 133
 - left, 127
 - projective, 146
 - quasi-projective, 149
 - right, 127
 - semisimple, 149
 - simple, 135
 - torsion, 170
 - torsion-free, 170
- Moore, E. H., 152, 166
- Moore, R., 340
- Moore–Penrose inverse, 152, 166–168
- Motzkin, T. S., 240, 276, 301, **307**, 310, 342
- Motzkin–Taussky theorem, 240, 255, 271, **271**, 274, 301, 310, 342, **342**
- multiplicity
 - algebraic, 6
 - geometric, 6
- Mumford, D., 340
- Murray, F. J., 197
- Neubauer, M. G., 118, 221, 224, 225, 240, 276, 297, 346
- Neubauer–Saltman theorem, 224, 225, 346
- Neubauer–Sethuraman theorem, 276, 297
- Nicholson, W. K., 4, 124
- nilpotency index, 7, 27
- nilpotent matrix, 7, 27
- Noether’s normalization theorem, 310, 326–328, 380
- Noether, E., 126, **382**
- Noetherian ring, 326
- nonhomogeneous
 - Weyr structure, 277
- nonvanishing set, 321
- nonderogatory matrix, 8, 103, 105
- norm of
 - matrix, 249
 - vector, 249
- null space, 5
- nullity
 - of matrix, 6
 - of transformation, 5
- Nullstellensatz
 - Hilbert’s, 326, 327, 350, 364
- O’Meara, K. C., 82, 153, 159, 160, 178, 189, 204, 239, 268, 276
- Omladič, M., 240, 307, 310, 364, 366, 368, 372, 377
- open
 - ball, 340
 - set, 321
- orthogonal
 - complement, 165
 - idempotents, 139
- orthonormal basis, 85
- Palin, M., xviii
- Pardo, E., 153, 159
- partition, 41
 - dual, 74
 - of n , 41
 - of a matrix, 12
- Passman, D. S., 173
- pattern frequency distribution, 245
- Pei, P., 40
- Penrose, R. A., 152, 166
- permutation matrix, 18, 48

- perturbation
 - 2-correctable, 262, 268
 - epsilon, 249
 - k -correctable, 261
- Phillips, H. B., 209
- phylogenetic invariant, 245
- PID, 170
- Plemmons, R. J., 238
- polynomial
 - characteristic, 6
 - map, 310, 314
 - minimal, 25, 354
 - symmetric, 318, 319
- primary decomposition, 29
- prime ideal, 350
- principal ideal domain, 170
- product of affine varieties, 335
- projection, 141, 167
- projective module, 146
- property
 - ASD, 310, 339
 - FGC, 173
- pseudo-inverse, 154
- pullback, 227
 - g -block, 227
- purely transcendental extension, 354

- quasi-finite map, 328
- quasi-inverse, 153
 - matrix, 268
- quasi-projective module, 149
- quaternions, 132, 237

- R -endomorphism, 133
- R -homomorphism, 133
- Rabinowitsch's trick, 350
- Rabinowitsch, S., 350
- radical
 - ideal, 350
 - of an ideal, 350
- rank
 - column, 6
 - full column-, 6
 - of transformation, 5
 - row, 6
- Raphael, R. M., 160, 178, 189
- rational
 - canonical form, 36
 - functions field, 354
 - map, 320
- Reams, R., 14
- reduced row-echelon, 35, 50
- reducible subset, 328
- reduction principle, 259
- regular
 - 1-, 8
 - 2-, 276
 - 3-, 230, 300
 - element, 152
 - k -, 8
 - ring, 153
- Reid, M., 326
- representations
 - of a group, 125
 - of a ring, 129
- Rhodes, J. A., 238, 243, 246, 249
- Richman, D. J., 62
- Riemann surface, 381
- right R -module, 127
- ring
 - co-ordinate, 347
 - exchange, 153
 - Noetherian, 326
 - regular, 153
 - von Neumann regular, 153
- ring-theoretic Krull dimension, 364
- Roby, Tom, 107
- Roman, S., xiv
- Rosenblum, M., 33
- Rosenthal, P., 33
- roundoff errors, 85
- row
 - equivalence, 35
 - rank, 6
- row-echelon form, 35

- $\mathcal{S}(3, n)$, 373
- Saltman, D., 118, 221, 224, 225, 346
- Schneider, H., 62, 81

- Schur, I., 32, 114
- Segre characteristic, 39
- self-centralizing subalgebra, 112, 221
- semisimple module, 149
- Semple, C., 243
- Separativity Problem, 159, 194
- Sergeichuk, V., 81
- Sethuraman, B. A., 223, 234, 240, 276, 297, 307
- Shafarevich, I. R., 359, 363
- Shapiro, H., 81
- Shekhtman, B., 239, 257
- similar
 - matrices, 18
 - unitarily, 85
- simple module, 135
- simultaneously
 - diagonalizable, **250**
 - triangularized, 70, **71**
- singular value decomposition, 166
- Šivic, K., 223, 234, 240, 307, 376
- Skolem–Noether theorem, 18
- $SL_n(F)$, 317, 320, 337, 358, 367
- Smith, S. P., 340, 378
- S_n , 18
- Song, Y. K., 267
- special linear group, 317
- Spiegel, E., 305
- splits, 147
- standard
 - basis, 5
 - form, 279
 - metric topology, 321
 - norm, 249
- Steel, M., 239, 243
- Strang, G., 62, 166, 167
- strictly block upper triangular, 14
- structure
 - homogeneous Weyr, 52
 - Jordan, **39**
 - Weyr, **50**
- subalgebra, 8
 - 1-generated, 206
 - 2-generated, 206
 - 3-generated commutative, 226
 - centralizer of, 221
 - generated by subset, 8
 - k -generated, 8, 203
 - maximal commutative, 112, 221
 - self-centralizing, 112, 221
- submodule, 132
 - cyclic, 134
 - finitely generated, 134
 - generated by, 134
 - sum, 134
- subset
 - closed, 320
 - dense, 329
 - irreducible, 328
 - open, 321
 - reducible, 328
- subspace
 - invariant, 19
 - leading edge, 109
- subspaces
 - direct sum of, 20
 - sum of, 5
- subvariety, 320
- sum
 - of a family of submodules, 149
 - of submodules, 134
 - of subspaces, 5
- summand
 - complementary, 141
 - direct, 141
- Sylvester's theorem, 33
- Sylvester, J., **43**, 236
- symmetric
 - group, 18
 - polynomial, 318, 319
- t -invariant subspaces, 133
- tableau
 - Young, 74, 107
- Taussky, O., 240, 276, 301, **308**, 310, 342
- Theorem
 - Burnside's on matrix subalgebras, 219
 - Burnside's orbit-counting, 108
 - Cayley's, 129
 - Cayley–Hamilton, 7, 236, 237
 - Chinese remainder, 210

- denseness, 340, 369, 378
- dimension of fibres, 363, 368
- f.g. torsion modules over a PID, 173
- fundamental homomorphism, 133
- Gerstenhaber's, 122, **219**, 225, 226, 234, 255, 273, **273**, 342, **343**
- Guralnick's, 364, **368**, 369, 370
- Hilbert's basis, 310, 326, 331
- Laffey–Lazarus, 224, 225, 346
- Lagrange's, 107
- Maschke's, 125
- Motzkin–Tausky, 240, 255, 271, **271**, 274, 301, 310, 342, **342**
- Neubauer–Saltman, 224, 225, 346
- Neubauer–Sethuraman, 276, 297
- Noether's normalization, 310, 326–328, 380
- Skolem–Noether, 18
- Sylvester's, 33
- Wedderburn–Artin, 6, 139, 258
- Toeplitz matrix, 68
- top row notation, 108
- topological Krull dimension, 363
- topology
 - Euclidean, 321, 339, **340**
 - induced, 328
 - metric, 321
 - standard metric, 321
 - Zariski, 310, 320
- torsion
 - element, 170
 - module, 170
- torsion-free
 - group, 128
 - module, 170
- tr deg, 355
- tr A , trace of A , 18
- trace map, 18, 317
- transcendancy
 - base, 354
 - degree, 354, 355
- transcendental element, 354
- translate, 231
- triangularized simultaneously, 70, **71**
- Trimble, H. C., 210
- Turnbull, H. W., 37
- $U(f)$, 321
- UFD, 172
- $U_k(\mathcal{A})$ leading edge subspace, 109
- unique factorization domain, 172
- unit-quasi-inverse, 154
- unit-regular, 195, 196
- unitarily similar, 85
- unitary
 - matrix, 85
 - transformation, 85
- University connection
 - Canterbury, 239
 - Connecticut, 82
 - Otago, 81
- $V(f)$, 311
- $V(S)$, 311
- varieties
 - isomorphic, 314
 - isomorphism of, 314
- variety
 - affine, 310, 311
 - irreducible, 310
- $[v]_{\mathcal{B}}$, 17
- vector
 - co-ordinate, 17
 - cyclic, 105
- Vinsonhaler, C., 82, 204, 239, 268, 276
- von Neumann
 - regular element, 152
 - algebra, 153, 197
 - regular ring, 153
- von Neumann, J., 152, **197**
- W -expansion, 278
- Ware, R., 157
- Warfield, R. B., Jr., 153
- Watanabe, J., 82
- Wedderburn–Artin theorem, 6, 139, 258
- Weyr
 - canonical form, 36, **61**
 - characteristic, 37, 61

- Weyr (*Continued*)
 decomposition of module, 184
 form, 36, **54**, 61
 algorithm, 82
 centralizer, **68**, 100
 for a nilpotent endomorphism, 177
 matrix, **54**
 basic, 49, **50**
 structure, **50**
 homogeneous, 52, 277
 nonhomogeneous, 277
 of general matrix, 61
Weyr, E., 37, 44, **94**
Wickless, W. J., 204
Wiles, A., 310
W-translate, 231
- Xue, W., 150
- Yakimov, M., 76
Young
 diagram, 75
 tableau, 74, 107
Yu, H.-P., 195
- Zariski
 closed subset, 321
 continuous, 325
 open subset, 321
 topology, 310, 320
Zariski, O., 320
Zhan, X., 40