TEA TIME LINEAR ALGEBRA

Explorations in Mathematics

2^{nd} edition

Leon Q. Brin

the second in a series of tea time textbooks

Southern Connecticut State University

To
Victorija, Cecelia, and Amy

Contents

Preface **ix**
About Tea Time Linear Algebra . . . ix
About the Exercises . . . xi
New in the Second Edition . . . xi
Acknowledgments . . . xi

I Matrix Mechanics

1 Matrix Calculations **3**
1.1 Matrices . . . 3
Key Concepts . . . 5
SageMath . . . 5
Exercises . . . 7
1.2 Component-wise Matrix Operations . . . 9
Key Concepts . . . 10
SageMath . . . 11
Exercises . . . 11
1.3 Matrix Multiplication . . . 13
Transposition and the Dot Product . . . 14
Key Concepts . . . 15
SageMath . . . 16
Exercises . . . 17
1.4 Magnitude and Orthogonality . . . 20
Key Concepts . . . 22
SageMath . . . 23
Exercises . . . 24
1.5 The Determinant . . . 27
Sudoku Row Linear Combinations . . . 28
Key Concepts . . . 30
SageMath . . . 30
Exercises . . . 31
1.6 Matrix "Division" . . . 33
A Formula for the Inverse (proven in section 3.7) . . . 35
One Property of the Inverse . . . 35
Inverses and Cryptography . . . 35
Key Concepts . . . 37
SageMath . . . 38
Exercises . . . 38

1.7 Eigenpairs . . . 42
Key Concepts . . . 44
SageMath . . . 44
Exercises . . . 45

2 Row Operations **49**
2.1 Systems of Linear Equations . . . 49
Elementary Matrices . . . 51
Key Concepts . . . 52
Exercises . . . 52
2.2 Row Reduction . . . 55
Key Concepts . . . 60
Exercises . . . 60
2.3 Existence, Uniqueness, and Echelon Forms . . . 62
SageMath . . . 66
Key Concepts . . . 67
Exercises . . . 67

3 Matrix Algebra **71**
3.1 Properties of Matrix Operations . . . 71
Background . . . 71
Some Properties of Matrices . . . 73
Applications to eigenpairs . . . 75
Key Concepts . . . 75
Exercises . . . 76
3.2 Matrix Equations . . . 78
Symbolic equations . . . 80
The most important equation in linear algebra . . . 80
Key Concepts . . . 82
Exercises . . . 82
3.3 Linear Independence . . . 85
Matrix Characterization Part 1 . . . 86
Key Concepts . . . 88
Exercises . . . 88
3.4 Characterization of $m \times n$ Matrices . . . 91
Key Concepts . . . 93
Exercises . . . 93
3.5 The Determinant Revisited . . . 95
Key Concepts . . . 103
Exercises . . . 103
3.6 Characterization of Square Matrices . . . 106
Key Concepts . . . 107
Exercises . . . 107
3.7 The Inverse Revisited . . . 110
Key Concepts . . . 112
Exercises . . . 112

II Matrix Abstraction

4 Vector Spaces and Inner Product Spaces **117**
4.1 Vector Spaces and Span . . . 117
Key Concepts . . . 119
Exercises . . . 120
4.2 Basis and Dimension . . . 125

Key Concepts . . . 127
Exercises . . . 127
4.3 Functions and Transformations . . . 130
Background . . . 130
Key Concepts . . . 132
Exercises . . . 134
4.4 Linear Transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ . . . 137
Key Concepts . . . 143
Exercises . . . 144
4.5 Isomorphisms . . . 147
Key Concepts . . . 149
Exercises . . . 149
4.6 Inner Product Spaces . . . 153
Key Concepts . . . 154
Exercises . . . 155

5 Exploring Vector Spaces and Inner Product Spaces **159**
5.1 Solution Spaces [3.3, 3.6, 4.1, 4.2] . . . 159
Key Concepts . . . 162
Exercises . . . 163
5.2 Coordinate Vectors [4.1, 4.2] . . . 167
Key Concepts . . . 169
Exercises . . . 170
5.3 Orthogonalization [4.6, 5.2] . . . 175
Key Concepts . . . 179
Exercises . . . 179
5.4 Similarity and Diagonalization [3.7, 4.4, 5.2] . . . 183
Similarity . . . 186
Key Concepts . . . 188
Exercises . . . 188

III Applications **193**

6 Mathematical Applications **195**
6.1 LU Factorization [3.4, 3.6] . . . 195
An Example . . . 197
Key Concepts . . . 199
Exercises . . . 199
6.2 The Power Method [3.5] . . . 201
Key Concepts . . . 205
Exercises . . . 205
6.3 Geometry: Determinants, Eigenvalues, and Area [3.3, 3.6, 4.4] . . . 209
Key Concepts . . . 213
Exercises . . . 214
6.4 Approximation [4.1, 4.6, 5.1, 5.3] . . . 217
Key Concepts . . . 221
Exercises . . . 221

7 Further Applications **227**
7.1 Linear Regression [6.4] . . . 227
Key Concepts . . . 233
Exercises . . . 233
7.2 Markov Chains [3.5] . . . 237
Formalities . . . 242

Key Concepts . . . 242
Exercises . . . 243
7.3 Fourier Series [4.6, 6.4, calculus] . . . 246
Key Concepts . . . 253
Exercises . . . 254
7.4 Discrete Dynamical Systems [3.2, 3.3] . . . 256
Key Concepts . . . 264
Exercises . . . 265
7.5 Rep-tiles [6.3] . . . 268
Key Concepts . . . 273
Exercises . . . 273

Solutions to Selected Exercises **279**

Answers to Selected Exercises **347**

Index **369**

Bibliography **373**

Preface

About Tea Time Linear Algebra

Greetings! And thanks for giving *Tea Time Linear Algebra* a read. The phrase "tea time" is meant to do more than give the book a catchy title. It is meant to describe the general nature of the discourse within. Much of the material will be presented as if it were being told to a student during tea time at University, but with the benefit of careful planning. There will be no big blue boxes highlighting the main points, no stream of examples after a short introduction to a topic, and no theorem... proof... theorem... proof structure. Instead, the necessary terms and definitions and theorems and examples will be woven into a more conversational style. My hope is that this blend of formal and informal mathematics will be easier to digest, and dare I say, students will be more invited to do their reading in this format.

Those who enjoy a more typical presentation might still find this textbook suits their needs. There will be a summary of the key concepts at the end of each conversation and a number of the exercises will be solved in complete detail in the appendix. One can get a closer-to-typical presentation by scanning for theorems in the conversations, reading the key concepts, and then skipping to the exercises with solutions. I hope most readers won't choose to do so, but it is an option. In any case, the exercises with solutions will be critical reading for most. Learning by example is often the most effective means. After reading a section, or at least scanning it, readers are strongly encouraged to skip to the statements of the exercises with solutions, contemplate their solutions, solve them if they can, and then turn to the back of the book for full disclosure. The hope is that, with their placement in the appendix, readers will be more apt to consider solving the exercises on their own before looking at the solutions.

The topical coverage in *Tea Time Linear Algebra* is fairly typical, but the order of presentation is not. The book starts with an introductory chapter covering all the typical matrix arithmetic including inverses and eigenvalues, but with only one method for each computation and without accompanying theory. The second chapter begins to bridge the gap between computational and theoretical linear algebra, covering row operations and systems of equations, concluding with the first theorem of the book, existence and uniqueness of solutions of linear systems. Chapter 3 introduces the notion of linear indepenence and revisits eigenpairs, inverses and determinants, adding depth to the computations of chapter 1. These three chapters conclude what might be considered the bare essentials of matrix algebra. Upon completion, students will be able to compute matrix sums and products, dot products, and lengths plus eigenpairs, matrix inverses, and determinants in multiple ways. They will be able to solve linear systems with any number of solutions and have enough theory to compute the number of solutions of a system without finding those solutions. This concludes part I, the mostly computational aspects of the course. Chapter 4 opens up part II by extolling the idea of abstraction. Vector spaces, basis, dimension, and isomorphisms are covered. Linear transformations and inner products are discussed for general vector spaces, not just $\mathbb{R}^n$. Chapter 5 closes part II by considering vector spaces such as column spaces, null spaces, and eigenspaces, and extending the ideas of orthogonality, length, and distance to arbitrary inner product spaces. The theme of abstraction is highlighted throughout. Part III (chapters 6 and 7) builds upon the computational aspects and theoretical notions of chapters 1-5 to solve mathematical and other problems, introducing unstudied theory of linear algebra sparingly. Factorization, iterative methods, geometry, and approximation take center stage. While these application sections largely stand on their own, the sections to which they refer are included parenthetically in the name of the section to help guide the reader and instructor on sequencing. Parts I and II do not have to be completed in their entirety before Part III is considered.

The first three chapters plus selections from chapters 4 and 5, capped with a smattering of chapters 6 and 7 cover what, at SCSU, constitutes a first semester course in linear algebra. It is likely full coverage of the book would require more than one semester. As this book is intended for use as a free download or an inexpensive print-on-demand volume, no effort has been made to keep the page count low or to spare copious diagrams and colors. In fact, I have taken the inexpensive mode of delivery as liberty to do quite the opposite. I have added many passages and diagrams that are not strictly necessary for the study of linear algebra, but are at least peripherally related, and may be of interest to some readers. Most of these passages will be presented as digressions, so they will be easy to identify. For example, the fact that determinants may be calculated by expansion along any row or any column is necessary basic fare for the course, but its proof is rather slippery and well beyond many students new to linear algebra. Its proof is therefore added as a “crumpet”. Other crumpets similarly cover technical details, but some lay out historical context and points for possible further study. They can be skipped without harm to the learning process, but are included to provide a more complete understanding of the fundamentals. In any case, each crumpet is there to enhance the reader’s understanding or appreciation of the subject, even if the material is not strictly necessary for an introductory study of linear algebra.

Many of the computations can not be done satisfactorily with pencil and paper, so sufficient linear algebra routines of SageMath are introduced and discussed. While one could simply ignore the SageMath sections and exercises and still get something out of this text, it is my firm belief that full appreciation for the content can not be achieved without getting one’s hands “dirty” by doing some calculation. It would be nice if readers have had at least some exposure to programming whether it be Python, Java, C, web programming, or just about anything, but I have made every effort to give enough detail so that even those who have never written a single line of code will be able to participate in this part of the study. In addition to maintaining a completely free learning experience, SageMath was chosen as the computer algebra system for this book because it allows linking to SageCells. Each live SageCell link in the PDF of this book leads to a bare bones, but fully operational portal to SageMath. Most links land on prepopulated code to aid with the process of using computer algebra. For example, in almost all cases the matrices involved in a question will be coded for the student to alleviate the tedium and errors of data entry. The SageMathCell website is offered freely to anyone and everyone!

As students come to linear algebra at widely varying levels of maturity, this course is not proof-based, nor does it require calculus. There are only 19 theorems and corollaries stated formally as such. Instead, main ideas and their proofs are often embedded in the course of discussion. Despite not being a proofs course, proofs are requested in the exercises, but usually using the word “justify” or “show”, which may be interpreted as requesting an informal argument for those who are not ready for full rigor. Almost never will the instructions begin “Prove…”, though students with rigorous proof experience are always most welcome to provide full rigor. In the end, the level of rigor is up to the reader and instructor. Several tips on proof technique, such as contraposition and induction, are sprinkled throughout the text to aid the unaccustomed reader in digesting some of the arguments, but the explanations are too scant to substitute as a complete course on foundations. References to calculus and exercises including integrals and derivatives can easily be ignored, with one exception. Section 7.3 on Fourier series necessarily requires calculus. Section 4.6 on inner product spaces is enhanced by knowledge of calculus, but does not require it. A corequisite course in calculus would suffice for section 4.6 but at least one complete semester is recommended for section 7.3.

Chapters 1 and 2 form the foundation for the rest of the text and every section therein should be covered in order before jumping to other topics. As an instructor, you may be tempted to fill in “missing” pieces of the discussion, but do whatever it takes to resist. The gaps will be filled in later. The purpose of these chapters is to give a straightforward introduction that gently eases the student into the finer details and provide a context for deeper study. While chapters 6 and 7 are placed at the back of the book, much of the material is appropriate long before the completion of the first 5 chapters. The applications within these final two chapters should be sprinkled into the course as prerequisite material is covered. For example, the first application, *LU* factorization, depends only on chapters 1-3 and can easily be included immediately following section 3.6, as indicated in the brackets of the section title. Bracketed lists of recommended prerequisite sections appear in the titles of all sections of chapters 5-7 to provide some guidance with sequencing. These prerequisite lists assume chapters 1 and 2 have been completed in their entirety, and are not meant to be hard and fast rules. You may find you are able to do without some of the recommendations, and you may be more comfortable adding others. The following table is included for further guidance.

interject this section	any time after completing this section
7.4 Discrete Dynamical Systems	3.4
7.2 Markov Chains; and 6.2 The Power Method	3.5
6.1 *LU* Factorization	3.6
6.3 Geometry; and 7.5 Rep-tiles	4.4

The remainder of the application sections (6.4, 7.1, 7.3) are best left until after the first 5 chapters, but you may find a way to modify the discussion of best approximation to avoid inner products and solution spaces, thus covering it sooner.

No matter how you choose to use the book, I hope you enjoy your reading of *Tea Time Linear Algebra*. It was my pleasure to write it. If you spot one of the many inaccuracies that have undoubtedly evaded my watch, please let me know. Feedback is always welcome.

Leon Q. Brin
brinl1@southernct.edu

About the Exercises

Exercises may be marked with one or more of the following symbols. In the PDF version of the book, each of these symbols is a hyperlink to the web or another part of the textbook. Click to follow.

[$\mathbb{S}$]-n	This exercise has a detailed solution on page n.
[$\mathbb{S}$]-n	Part of this exercise has a detailed solution on page n.
[$\mathbb{A}$]-n	This exercise has an answer on page n.
[$\mathbb{A}$]-n	Part of this exercise has an answer on page n.
SageMathCell	SageMath is recommended or required for this exercise.
	GeoGebra is recommended or required for this exercise.

Answers and solutions may be followed back to the exercise by clicking the exercise number.

New in the Second Edition

The biggest difference between the first and second editions is the launch of an accompanying MyOpenMath course shell. Instructors can now assign and collect homework perfectly aligned with the textbook online. Create randomly generated questions for assessments. Communicate with the class. Maintain an online gradebook. Post notes to the online calendar. Full course management available. Most of the questions in the text have an online cousin, and most of the cousins contain randomly generated numbers. There are a few questions in the book not available online, and there are a few questions available online that are not in the book.

The online course contains prepopulated assignments for each section including the reading for the section, instructor's notes, and a limited number of links to additional resources, all of which can easily be modified to suit your needs. To get started, request an instructor's account at MyOpenMath today. It's free to sign up and free to use. No charge ever. Once you have your account, you will find the Tea Time Linear Algebra course in the Course Browser at MyOpenMath.

Acknowledgments

I gratefully acknowledge the generous support I received during the writing of this textbook. From the patience my immediate family, Amy, Cecelia, and Victorija exercised while I was absorbed by my laptop's screen to the time allowed by sabbatical leave to the numerous flaws pointed out by students, this book is not a one-man show despite the single credit.

Part I

Matrix Mechanics

Chapter 1

Matrix Calculations

1.1 Matrices

The fundamental object of linear algebra is the **matrix**. A matrix is very much like a table or a spreadsheet, but without headings, labels, or lines. The data in a matrix are separated by space. The whole matrix is enclosed by large parentheses or square brackets, but is otherwise unadorned.

Crumpet 1: Dictionary Definition

ma•trix (mā′ trĭks) *n.*, *pl.* **ma•tri•ces** (mā′ trĭ-sēz′). *Math.* A rectangular array of algebraic quantities *usu.* delimited by parentheses or square brackets.

The following are all matrices.

$$\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \quad \begin{pmatrix} 3 & -4 & 63 \\ 2 & 8 & -17 \end{pmatrix} \quad \begin{bmatrix} \hat{i} & \hat{j} & \hat{k} \\ -1 & 6 & \sqrt{3} \\ \sqrt{2} & 13 & e \end{bmatrix}$$

$$\begin{pmatrix} .929 & .988 & .405 & .877 & .752 & .541 & .269 \\ .390 & .595 & .186 & .328 & .315 & .566 & .478 \\ .731 & .224 & .254 & .543 & .575 & .499 & .881 \end{pmatrix}$$

The **size of a matrix** is described by its number of rows "by" its number of columns, and is abbreviated as in 2×3, read "two by three". A 7×5 matrix has seven rows and 5 columns. The number of rows is always listed first. The rows are indexed from top to bottom, and the columns are indexed from left to right. The first row is the topmost row, and the first column is the leftmost column. There are no restrictions on the numbers of rows or columns other than each must be a positive integer. The individual quantities in a matrix are called **entries**. The entry in the i^{th} row (from the top) and j^{th} column (from the left) of a matrix is called the i,j-entry. The row number always precedes the column number.

Matrices are most often labeled by capital letter variables such as A, B, or M. This helps distinguish them from numerical variables such as x, y, z, s, or t. In this text, the i,j-entry of a matrix A is denoted $A_{i,j}$. The 5,1-entry (fifth row, first column) of a matrix M is denoted $M_{5,1}$, for example.

Crumpet 2: Other Notations for Entries

The subscripted lower case counterpart to a matrix variable is often used to represent the entries of a matrix. You will often see $b_{1,2}$ or even b_{12} represent the 1,2-entry of B. Don't be surprised when you run into it!

Two matrices are **equal** if they are the same size and corresponding entries are equal.

Taking a cue from computer science and the currently popular Python programming language, the i^{th} row of a matrix B is denoted $B_{i,:}$, the : indicating that all columns of the row are included. The j^{th} column of the same matrix B is denoted $B_{:,j}$, where the : indicates that all rows are included.

$$\text{If } B = \begin{bmatrix} 2 & 6 & 1 & 8 \\ -3 & 4 & -2 & 1 \\ -2 & -5 & 4 & 1 \end{bmatrix} \text{ then } B_{2,:} = \begin{bmatrix} -3 & 4 & -2 & 1 \end{bmatrix} \text{ and } B_{:,4} = \begin{bmatrix} 8 \\ 1 \\ 1 \end{bmatrix}$$

A **submatrix** of a matrix M is any matrix derived by deleting some number of rows (less than the total number of rows) and some number of columns (less than the total number of columns) from M.

$$\begin{bmatrix} 2 & 6 & 1 \\ -3 & 4 & -2 \end{bmatrix} \text{ is a submatrix of } \begin{bmatrix} 2 & 6 & 1 & 8 \\ -3 & 4 & -2 & 1 \\ -2 & -5 & 4 & 1 \end{bmatrix}$$

derived by deleting the last row and last column. $\begin{bmatrix} 6 \end{bmatrix}$ is a submatrix of $\begin{bmatrix} 2 & 6 & 1 \\ -3 & 4 & -2 \end{bmatrix}$, derived by deleting the second row, the first column, and the third column. A submatrix derived by deleting one row and one column of a matrix is common enough that we use a special notation for it: $B_{\backslash i,j}$ (read "B without row i and column j").

$$\text{If } B = \begin{bmatrix} 2 & 6 & 1 & 8 \\ -3 & 4 & -2 & 1 \\ -2 & -5 & 4 & 1 \end{bmatrix} \text{ then } B_{\backslash 2,3} = \begin{bmatrix} 2 & 6 & 8 \\ -2 & -5 & 1 \end{bmatrix}.$$

Though we will not make frequent use of it, the : notation can be used to identify submatrices other than single columns or single rows by placing a number before and a number after the colon as in $2:5$, which means rows (or columns) two through five. For example, $B_{1:2,1:3}$ represents the submatrix of B consisting of its first two rows and first three columns. All other rows and columns are excluded. $B_{2:7,3}$ represents the submatrix of B consisting of rows two through seven of column three.

Often the entries of a matrix will have underlying meaning, derived from the context of a story problem or application. A table or spreadsheet of common grocery items at various stores such as

Price Comparison ($)

		Item		
		eggs	milk	bananas
Store	Food Plus	2.89	4.69	2.07
	Grocer Girl	3.69	4.99	2.37
	Eddie's Eats	2.79	4.29	2.57

or

	A	B	C	D	E
1		**Price Comparison ($)**			
2			eggs	milk	bananas
3	**Store**	Food Plus	2.89	4.69	2.07
4		Grocer Girl	3.69	4.99	2.37
5		Eddie's Eats	2.79	4.29	2.57

would be summarized in a matrix as

$$\begin{bmatrix} 2.89 & 4.69 & 2.07 \\ 3.69 & 4.99 & 2.37 \\ 2.79 & 4.29 & 2.57 \end{bmatrix}.$$

All the descriptive words are stripped. Labels would only get in the way of any mathematical operation, so the rows and columns of a matrix are not labeled. Their meaning must be communicated some other way. View this video[1] (3:13) for more examples where a matrix might be useful.

While the meaning of the entries in the grocery items example is stated explicitly in the table/spreadsheet, there are times when meaning will simply be implied or understood from context. In any case, if the numbers in a matrix are to have contextual meaning, that information must be supplied separately.

[1] https://youtu.be/BZWFkUQ3tco?t=71

Key Concepts

matrix A rectangular array of algebraic quantities usually delimited by parentheses or square brackets. Upper case letters are used for variables representing matrices.

(matrix) entry One of the individual quantities in a matrix.

(matrix) size The number of rows "by" the number of columns.

matrix equality Two matrices are equal if they are the same size and corresponding entries are equal.

submatrix The matrix resulting from deleting some number of rows (less than the total number of rows) and some number of columns (less than the total number of columns) from a matrix.

notation $A, B, \ldots, M, \ldots$ Upper case letters are used for variables representing matrices.

$A_{i,j}$ The entry in row i and column j of matrix A.

$A_{m,:}$ Row m of matrix A.

$A_{:,n}$ Column n of matrix A.

$A_{\setminus m,n}$ The submatrix of A consisting of all entries except those in row m or in column n.

$A_{i:j,k:l}$ The submatrix of A consisting of rows i through j of columns k through l.

$m \times n$ The size of a matrix with m rows and n columns.

SageMath

The matrices and operations of this section (and the entire text) can be handled electronically by SageMath. All you need is the syntax, the proper combinations of words and symbols. A matrix must be defined before it can be used in any computation. In SageMath, there are several ways to define a matrix, but we will most often use the syntax

```
M = matrix(rows,cols,[list of entries])
```

The `rows` and `cols` stand for the number of rows and number of columns in the matrix, respectively. The `list of entries` must be comma-separated as in `1,2,3`. Entered into SageMath properly, this line creates the variable `M`

Crumpet 3: Instantiation

In computer science, giving a value to a variable is called *instantiation*.

as a matrix from which we can extract entries or submatrices, print out, or perform operations. Just as we can on paper, we can name the matrix using any letter. It does not have to be `M`.

Submatrices can be extracted using : notation just as we have been doing on paper. In SageMath, though, subscripts aren't used. Square brackets are. So $M_{3,:}$ would be written `M[2,:]` in SageMath. Yes, that looks like a typo. It is not! On paper, and in mathematics generally, we index the rows and columns of matrices in a way that seems most natural. The first row is row 1, the second row is row 2, and so on. However, SageMath uses the very common computer programming convention of 0-indexing. Counting starts with 0 instead of 1 in SageMath. So the first row of a matrix `M` (in SageMath, Python, and many other programming languages) is row 0, the second row is row 1, and so on.

The square bracket notation is used to extract entries of a matrix, too. In SageMath the i,j-entry of a matrix `M` is indicated by `M[i-1,j-1]`. Table 1.1 summarizes the extraction of entries and submatrices using SageMath.

In SageMath, the lines

```
M = matrix(3,2,[1,2,3,4,5,6])
S = M[2,:]
```

Table 1.1: Matrices, entries, and submatrices in SageMath.

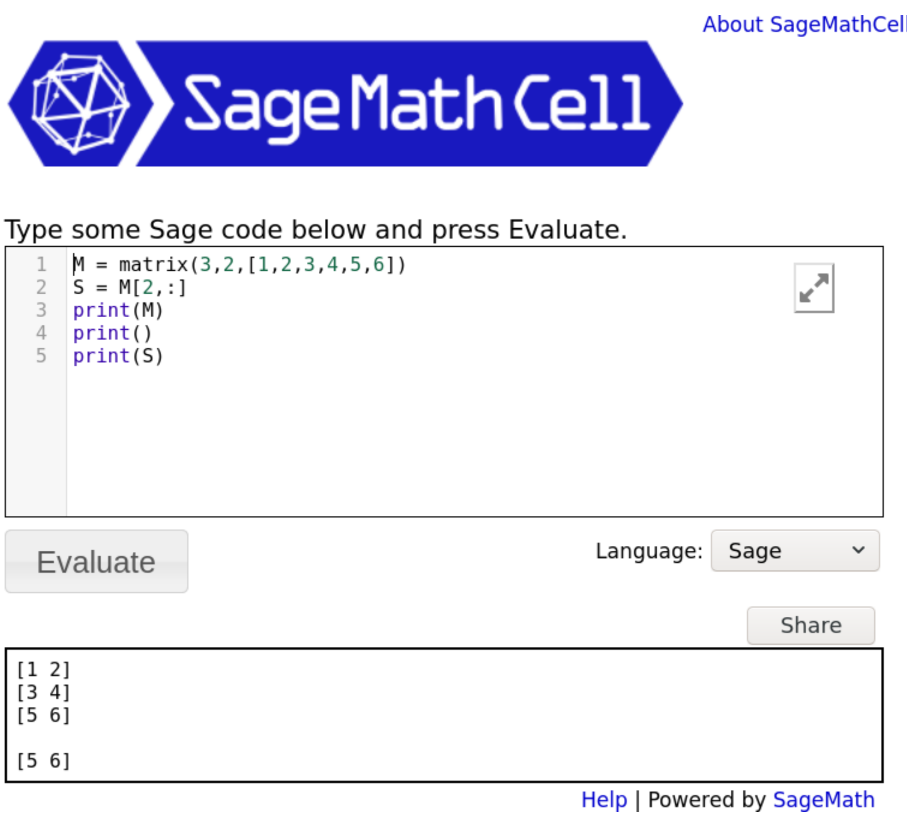
	Mathematics	SageMath (0-indexed)
matrix	$M = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$	`M = matrix(3,2,[1,2,3,4,5,6])`
row	$M_{r,:}$	`M[r-1,:]`
column	$M_{:,c}$	`M[:,c-1]`
submatrix	$M_{i:j,k:l}$	`M[i-1:j,k-1:l]`
submatrix	$M_{\backslash r,c}$	`M.delete_rows([r-1]).delete_columns([c-1])`
entry	$M_{r,c}$	`M[r-1,c-1]`

define a matrix `M` and a submatrix `S`, but do not create any output. When the code is run (evaluated), it seems nothing has happened! Test it by following this link: 1. Rest assured, these lines cause SageMath to do things internally. We just aren't seeing the results yet.

If we add a couple lines requesting the display of our matrices, we will see the results. In SageMath, this can be done with the `print(object)` statement. In this case, we want to print out two matrices. A little space between them would be good too. Printing "nothing", using `print()`, actually prints a blank line. The following SageMath code creates a matrix `M`, a submatrix `S`, and prints them both with a blank line between.

```
M = matrix(3,2,[1,2,3,4,5,6])
S = M[2,:]
print(M)
print()
print(S)
```

Here is a screenshot of this code being processed at `SageCell.SageMath.org`.



Live at SageMathCell 2.

Crumpet 4: Nested Statements in SageMath

SageMath statements may be nested. One statement may appear as the argument (inside) of another. For example, the code

```
M = matrix(3,2,[1,2,3,4,5,6])
S = M[2,:]
print(M)
print()
print(S)
```

might also be written as follows.

```
M = matrix(3,2,[1,2,3,4,5,6])
print(M)
print()
print(M[2,:])
```

Notice the extraction of the third row of `M` happens inside the `print` statement. There is no need to produce a variable named `S` since it is not used for any other purpose.

Exercises

1. How many rows does a matrix with the given size have?
   (a) 15×6 [𝕊]-279
   (b) 6×8
   (c) 1×11
   (d) 17×2
2. How many columns does a matrix with the given size have?
   (a) 5×10
   (b) 12×5 [𝕊]-279
   (c) 6×12
   (d) 18×19
3. How many entries does a matrix with the given size have?
   (a) 3×13
   (b) 9×8
   (c) 4×14 [𝕊]-279
   (d) 7×6
4. Identify the requested entry of the given matrix.
   (a) $A_{2,4}$
   $$A = \begin{bmatrix} 23 & 31 & 44 & -9 & 45 \\ 27 & -6 & 14 & 33 & -33 \\ -22 & 48 & -17 & -48 & 41 \end{bmatrix}$$
   (b) $B_{1,2}$
   $$B = \begin{bmatrix} -3 & 39 & -1 \\ 3 & -30 & 7 \\ -27 & -48 & 32 \end{bmatrix}$$
   (c) $P_{3,4}$
   $$P = \begin{bmatrix} 47 & 14 & -10 & 10 & -11 \\ 21 & -29 & -39 & 49 & -26 \\ -22 & 20 & 12 & 37 & 44 \\ -18 & -37 & -30 & -42 & -17 \end{bmatrix}$$
   (d) $M_{3,1}$
   $$M = \begin{bmatrix} 21 & -14 & 43 & 34 \\ 8 & -32 & -3 & -20 \\ -2 & 50 & -24 & 20 \end{bmatrix}$$ [𝕊]-279
5. Let
   $$N = \begin{bmatrix} -11 & -2 & -6 & -4 & -3 & 5 \\ -5 & 12 & 3 & -2 & -4 & -7 \\ -8 & -12 & 11 & -12 & 4 & -3 \\ 3 & 10 & 9 & 0 & -7 & -10 \\ -2 & 12 & -9 & 3 & -5 & 8 \end{bmatrix}.$$ What is the size of the submatrix?
   (a) $N_{5,:}$
   (b) $N_{1,:}$ [𝔸]-347
   (c) $N_{:,2}$
   (d) $N_{:,3}$ [𝕊]-279
   (e) $N_{1:5,2:4}$
   (f) $N_{2:3,4:5}$ [𝔸]-347
   (g) $N_{\backslash 3,5}$
   (h) $N_{\backslash 4,2}$ [𝕊]-279

6. Let $A = \begin{bmatrix} -7 & 2 & 1 & 8 & -1 \\ -8 & -11 & 10 & -6 & 1 \\ 1 & -10 & 12 & -12 & 0 \\ 9 & -1 & 3 & -6 & -8 \\ 6 & -9 & 3 & 4 & -5 \\ 2 & -4 & 10 & -7 & -3 \end{bmatrix}$. Identify the submatrix of A.

   (a) $A_{3,:}$

   (b) $A_{6,:}$ [$\mathbb{S}$]-279

   (c) $A_{:,2}$

   (d) $A_{:,5}$ [$\mathbb{A}$]-347

   (e) $A_{2:3,3:5}$

   (f) $A_{2:4,2:3}$ [$\mathbb{S}$]-279

   (g) $A_{\backslash 4,2}$

   (h) $A_{\backslash 3,5}$ [$\mathbb{A}$]-347

7. Let $P = \begin{bmatrix} 11 & 7 & 4 \\ -8 & -6 & 1 \\ -2 & -1 & 5 \\ 8 & 3 & 12 \end{bmatrix}$. Supply notation for the entry or submatrix.

   (a) 5

   (b) $\begin{bmatrix} 11 & 4 \\ -8 & 1 \\ -2 & 5 \end{bmatrix}$

   (c) $\begin{bmatrix} -8 & -6 & 1 \\ -2 & -1 & 5 \end{bmatrix}$

   (d) $\begin{bmatrix} 8 & 3 & 12 \end{bmatrix}$

   (e) -2

   (f) $\begin{bmatrix} 11 \\ -8 \\ -2 \\ 8 \end{bmatrix}$

8. SageMathCell Let $B = \begin{bmatrix} 6 & -7 & 1 \\ 11 & -11 & -2 \\ 11 & 12 & 12 \\ -7 & 9 & 10 \end{bmatrix}$. Write SageMath code that will accomplish the following.

   (a) Create the matrix B.

   (b) Print B.

   (c) Extract $B_{2,3}$.

   (d) Print $B_{2,3}$.

9. SageMathCell Let $C = \begin{bmatrix} -4 & -9 & 13 & -11 \\ 7 & 5 & -14 & -2 \\ -12 & 12 & 8 & 11 \end{bmatrix}$. Write SageMath code that will accomplish the following. [$\mathbb{A}$]-347

   (a) Create the matrix C.

   (b) Print C.

   (c) Extract $C_{3,2}$.

   (d) Print $C_{3,2}$.

10. SageMathCell 3 Add code that will print out (i) the third row and (ii) the first column of D.

```
D = matrix(3,4,[-2,10,-7,8,-2,7,
                11,-7,-4,5,6,10])
```

    What is the output of your code?

11. SageMathCell 4 If you swap the 3 and the 4 in the code of exercise 10, as shown below, what is the new output of your code (third row and first column of D)?

```
D = matrix(4,3,[-2,10,-7,8,-2,7,
                11,-7,-4,5,6,10])
```

1.2 Component-wise Matrix Operations

While a sudoku board is not a matrix, if we strip away the color and the lines, it certainly is a rectangular array of numbers, the essence of a matrix. Soon we will do just that, but for now let's have a look at the sudoku board without thinking about matrices. Notice it consists of nine 3×3 blocks.

1	4	7	8	6	5	2	3	9
2	6	3	1	9	4	5	7	8
8	5	9	3	7	2	1	6	4
3	2	1	7	4	6	8	9	5
9	7	5	2	8	3	4	1	6
4	8	6	9	5	1	3	2	7
6	3	4	5	2	7	9	8	1
5	9	2	6	1	8	7	4	3
7	1	8	4	3	9	6	5	2

Pick your favorite two 3×3 blocks and think about how you might add them to one another. Don't just read on. Stop and think about this briefly. There are no right or wrong answers. What would be one reasonable way to add two blocks? If you are like most students, you probably came up with one of two ways to add the blocks. The first one is to add all the numbers in each block. If you did this, you should have gotten 90 for the total. Sum the numbers in a different pair of blocks, and you will notice you get 90 again. The sum of the 18 numbers in any two blocks is 90. Can you see why? Answer on page 12. This way of adding is legitimate, but maybe a little unsatisfying since the sum is always 90.

What if the sum of the two blocks were another 3×3 block? This way of thinking has a lot of precedent in mathematics. The sum of two integers is an integer. The sum of two rational numbers is a rational number. The sum of two functions is a function. The sum of two areas is an area. The operation of addition always seems to preserve the type of object being added.

Crumpet 5: Operators

In mathematics a **binary operator**, such as +, takes two objects (inputs or addends) from a set and produces a third object (output or sum) from the same set.

With this idea in hand, perhaps the most organized way to proceed is to add the number in the upper-left corner of the first block to the number in the upper-left corner of the second block to produce the number in the upper-left corner of the sum. Similarly, the other 8 numbers of the sum can be produced by adding corresponding numbers (by location) of the two blocks being added. Here is an illustration of that process.

$$\begin{array}{|c|c|c|}\hline 1&4&7\\\hline 2&6&3\\\hline 8&5&9\\\hline\end{array} + \begin{array}{|c|c|c|}\hline 7&4&6\\\hline 2&8&3\\\hline 9&5&1\\\hline\end{array} = \begin{array}{|c|c|c|}\hline 1+7&4+4&7+6\\\hline 2+2&6+8&3+3\\\hline 8+9&5+5&9+1\\\hline\end{array} = \begin{array}{|c|c|c|}\hline 8&8&13\\\hline 4&14&6\\\hline 17&10&10\\\hline\end{array}$$

The exact same component-wise (entry-by-entry) mechanics are used for adding matrices. Using matrix entry notation,

$$\text{if } A \text{ and } B \text{ are matrices, then } (A+B)_{i,j} = A_{i,j} + B_{i,j}$$

for all entries $A_{i,j}$ and $B_{i,j}$ of A and B. That is, the i,j-entry of $A+B$ is the sum of the i,j-entries of A and B. For the sum to be defined "for all entries $A_{i,j}$ and $B_{i,j}$" A and B must have the exact same size. The sum of matrices of differing size is undefined. Subtraction of matrices is defined analogously.

$$\text{If } A \text{ and } B \text{ are matrices, then } (A-B)_{i,j} = A_{i,j} - B_{i,j}$$

for all entries $A_{i,j}$ and $B_{i,j}$ of A and B. The difference of matrices of differing size is undefined.

All that is to say we add matrices the same way we added the sudoku blocks and we can subtract matrices in a similar manner. Transferring the numbers of a sudoku board to a matrix is good practice in creating matrices where there are none, extracting them from their context for mathematical work. Let's start by looking at each 3×3 block of the sudoku board as a matrix.

1	4	7	8	6	5	2	3	9
2	6	3	1	9	4	5	7	8
8	5	9	3	7	2	1	6	4
3	2	1	7	4	6	8	9	5
9	7	5	2	8	3	4	1	6
4	8	6	9	5	1	3	2	7
6	3	4	5	2	7	9	8	1
5	9	2	6	1	8	7	4	3
7	1	8	4	3	9	6	5	2

$$\Rightarrow \quad \begin{matrix} \begin{bmatrix} 1 & 4 & 7 \\ 2 & 6 & 3 \\ 8 & 5 & 9 \end{bmatrix} & \begin{bmatrix} 8 & 6 & 5 \\ 1 & 9 & 4 \\ 3 & 7 & 2 \end{bmatrix} & \begin{bmatrix} 2 & 3 & 9 \\ 5 & 7 & 8 \\ 1 & 6 & 4 \end{bmatrix} \\ \begin{bmatrix} 3 & 2 & 1 \\ 9 & 7 & 5 \\ 4 & 8 & 6 \end{bmatrix} & \begin{bmatrix} 7 & 4 & 6 \\ 2 & 8 & 3 \\ 9 & 5 & 1 \end{bmatrix} & \begin{bmatrix} 8 & 9 & 5 \\ 4 & 1 & 6 \\ 3 & 2 & 7 \end{bmatrix} \\ \begin{bmatrix} 6 & 3 & 4 \\ 5 & 9 & 2 \\ 7 & 1 & 8 \end{bmatrix} & \begin{bmatrix} 5 & 2 & 7 \\ 6 & 1 & 8 \\ 4 & 3 & 9 \end{bmatrix} & \begin{bmatrix} 9 & 8 & 1 \\ 7 & 4 & 3 \\ 6 & 5 & 2 \end{bmatrix} \end{matrix}$$

Previously we added the upper-left block and the middle block of the sudoku board. Now let's add the upper-left matrix and the middle matrix:

$$\begin{bmatrix} 1 & 4 & 7 \\ 2 & 6 & 3 \\ 8 & 5 & 9 \end{bmatrix} + \begin{bmatrix} 7 & 4 & 6 \\ 2 & 8 & 3 \\ 9 & 5 & 1 \end{bmatrix} = \begin{bmatrix} 1+7 & 4+4 & 7+6 \\ 2+2 & 6+8 & 3+3 \\ 8+9 & 5+5 & 9+1 \end{bmatrix} = \begin{bmatrix} 8 & 8 & 13 \\ 4 & 14 & 6 \\ 17 & 10 & 10 \end{bmatrix}$$

Conceptually, it is the same computation.

Multiplying a matrix by a number is also done component-wise. Multiplying the bottom-left 3×3 matrix extracted from our sudoku board by 5 is done as follows.

$$5\begin{bmatrix} 6 & 3 & 4 \\ 5 & 9 & 2 \\ 7 & 1 & 8 \end{bmatrix} = \begin{bmatrix} 5\cdot6 & 5\cdot3 & 5\cdot4 \\ 5\cdot5 & 5\cdot9 & 5\cdot2 \\ 5\cdot7 & 5\cdot1 & 5\cdot8 \end{bmatrix} = \begin{bmatrix} 30 & 15 & 20 \\ 25 & 45 & 10 \\ 35 & 5 & 40 \end{bmatrix}$$

This is often referred to as scalar[2] multiplication to differentiate it from matrix multiplication, the subject of the next section. In symbols,

$$\text{If } A \text{ is a matrix and } c \text{ is a scalar, then } (cA)_{i,j} = cA_{i,j}$$

for all entries $A_{i,j}$ of A. This means that cA has the same size as A and the i,j-entry of cA is c times the i,j-entry of A. To be complete Ac is defined to equal cA.

Crumpet 6: Fields

Sets of scalars other than real numbers and complex numbers are permissible in linear algebra as long as matrix entries come from the same field. A field must contain an additive identity, denoted 0, and a multiplicative identity, denoted 1. A field with only these two elements can be defined by treating 0 and 1 as integers except that $1 + 1 = 0$. The field of two elements is often denoted $\mathbb{F}_2$ or $\mathbb{Z}_2$.

Key Concepts

binary operator A function with two inputs and one output, all three from the same set.

[2] In this textbook, the word *scalar* refers to either a real number or a complex number. In more abstract settings, the word scalar refers to any element of a field.

matrix addition For any matrices A and B of the same size, the sum $A+B$ is defined, has the same size as A and B, and $(A+B)_{i,j} = A_{i,j} + B_{i,j}$ for all entries $A_{i,j}$ and $B_{i,j}$. If A and B differ in size, then $A+B$ is undefined.

matrix subtraction For any matrices A and B of the same size, the difference $A-B$ is defined, has the same size as A and B, and $(A-B)_{i,j} = A_{i,j} - B_{i,j}$ for all entries $A_{i,j}$ and $B_{i,j}$. If A and B differ in size, then $A-B$ is undefined.

scalar An element of a field.

scalar multiplication For any matrix A and scalar c, the scalar product cA is defined, has the same size as A, and $(cA)_{i,j} = cA_{i,j}$ for all entries $A_{i,j}$. Moreover, Ac is defined to equal cA.

SageMath

The syntax for scalar multiplication, matrix addition, and matrix subtraction in SageMath is much like calculator syntax. The plus sign is used for addtion, the minus sign for subtraction, and the asterisk for multiplication. The asterisk is not optional. Typing two quantities with no operator between produces an error. Multiplication is not implied by lack of a symbol. SageMath code that reproduces the calculations of this section follows.

```
A=matrix(3,3,[1,4,7,2,6,3,8,5,9])
B=matrix(3,3,[7,4,6,2,8,3,9,5,1])
print(A+B)
print()
C=matrix(3,3,[6,3,4,5,9,2,7,1,8])
print(5*C)
```

SageMathCell [5]. The output is as follows.

```
[ 8  8 13]
[ 4 14  6]
[17 10 10]

[30 15 20]
[25 45 10]
[35  5 40]
```

Exercises

1. Perform the operation if possible.

(a) $\begin{bmatrix} -1 & -6 & 0 \\ -6 & -5 & 10 \end{bmatrix} + \begin{bmatrix} 1 & -10 & 3 \\ 9 & 0 & 2 \end{bmatrix}$

(b) $\begin{bmatrix} 1.6 & 8.4 \\ 8.16 & -0.33 \end{bmatrix} + \begin{bmatrix} 4.01 & 1.75 \\ 9.35 & 1.49 \\ -0.24 & 0.58 \end{bmatrix}$

(c) $\begin{bmatrix} -5 & -8 & 7 & 5 \\ -9 & -3 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 3 & -4 & 7 & -8 \\ 1 & -2 & 2 & -5 \end{bmatrix}$ [$\mathbb{A}$]-347

(d) $\begin{bmatrix} 4 & 1 \\ 3 & 2 \end{bmatrix} - \begin{bmatrix} -10 & -3 \\ 1 & 8 \end{bmatrix}$

(e) $\begin{bmatrix} -6 \\ 0 \\ -6 \end{bmatrix} + \begin{bmatrix} 9 \\ 10 \\ 0 \end{bmatrix}$

(f) $2\begin{bmatrix} 5 & -11 & -2 \\ 14 & 1 & -8 \\ 13 & -1 & 6 \end{bmatrix}$

(g) $\begin{bmatrix} 3.43 & 6.59 \\ -0.96 & 0.16 \end{bmatrix} + \begin{bmatrix} -0.78 & 8.68 \\ 2.14 & 8.79 \end{bmatrix}$

(h) $\begin{bmatrix} -9 & 1 & 5 \\ 10 & 1 & -10 \\ 2 & -3 & -7 \end{bmatrix} + \begin{bmatrix} -2 & 7 & 8 \\ 9 & -4 & 9 \\ 2 & -10 & -10 \end{bmatrix}$

(i) $2\begin{bmatrix} -1 & 6 \\ 8 & 15 \end{bmatrix}$ [$\mathbb{S}$]-279

(j) $\begin{bmatrix} 4.65 & 1.33 & 8.86 \\ 6.03 & 4.56 & 4.8 \end{bmatrix} - \begin{bmatrix} 1.85 & 6.4 & 7.33 \\ 4.58 & 8.39 & 1.89 \end{bmatrix}$

(k) $\begin{bmatrix} 0 & 6 & -8 & -2 \\ 8 & 10 & 7 & -3 \end{bmatrix} + \begin{bmatrix} 9 & 2 \\ 6 & -3 \end{bmatrix}$

(l) $\begin{bmatrix} 4.83 & 7.65 \\ -0.48 & 7.82 \\ 0.25 & 2.53 \end{bmatrix} - \begin{bmatrix} 4.44 & 6.57 \\ 4.22 & 7.17 \end{bmatrix}$

(m) $\begin{bmatrix} 1 & -9 & 6 & 10 \end{bmatrix} - \begin{bmatrix} -2 & -1 & 2 & -7 \end{bmatrix}$

(n) $\begin{bmatrix} 9 & -10 & 4 \\ 10 & 10 & 1 \\ 0 & -8 & 3 \end{bmatrix} - \begin{bmatrix} 10 & -4 & -4 \\ 8 & 9 & 4 \end{bmatrix}$ [$\mathbb{S}$]-279

(o) $\begin{bmatrix} 2 \\ 7 \\ 9 \\ 3 \end{bmatrix} + \begin{bmatrix} -7 & 8 \\ 3 & 9 \\ -4 & 2 \\ 2 & 9 \end{bmatrix}$

(p) $\begin{bmatrix} 10 \\ 4 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 & 10 & -2 \end{bmatrix}$

(q) $3.19 \begin{bmatrix} -12.96 \\ -0.96 \\ -7.99 \\ 11.05 \end{bmatrix}$

(r) $\begin{bmatrix} -6 & -4 \\ 7 & -2 \\ -1 & 0 \end{bmatrix} - \begin{bmatrix} -4 & -6 \\ 10 & 5 \\ -3 & 8 \end{bmatrix}$

2. Suppose M is a 5×5 matrix and $M + N$ is defined (the sum can be computed). How many entries does N have?

3. In your own words, describe how to add or subtract two matrices, and explain how to determine whether the addition or subtraction can be done.

4. Can a matrix with 29 nonzero entries be added to a matrix with 25 nonzero entries? Explain. [𝔸]-347

5. Suppose M and N are matrices such that their sum is defined ($M + N$ can be computed). Is the following true or false? Explain.

$$M + N = N + M$$

6. Suppose M and N are matrices such that their difference is defined ($M - N$ can be computed). Is the following true or false? Explain.

$$M - N = N - M$$

[𝕊]-279

7. Suppose M is a matrix of size 3×7, c is a scalar, and the matrix computation cM is defined. What is the size of matrix cM?

For the remaining exercises, let

$$A = \begin{bmatrix} 42 & 0 & -47 & -34 & -10 & -48 \\ 8 & 26 & 43 & -18 & -20 & -30 \\ -41 & -40 & -29 & -36 & -44 & 12 \\ -42 & 47 & 28 & 4 & 38 & -22 \\ 18 & -15 & -1 & 29 & 37 & 9 \end{bmatrix} \quad N = \begin{bmatrix} -21 & -33 & 28 & -15 & 34 & 45 \\ 27 & 40 & -13 & -23 & -10 & 15 \\ 43 & -6 & 46 & 17 & 13 & 21 \\ -40 & -46 & 2 & 16 & 22 & -14 \\ 10 & -12 & 29 & 35 & 48 & -31 \end{bmatrix}$$

$$Q = \begin{bmatrix} -17 & -37 & -34 & 20 & -14 & 10 \\ -23 & 44 & 47 & 18 & 19 & 49 \\ 11 & 33 & 35 & -50 & 2 & 9 \\ -36 & -18 & 7 & 17 & -49 & 31 \\ -8 & 16 & 28 & -32 & -2 & 5 \end{bmatrix} \quad T = \begin{bmatrix} 40 & 47 & 13 & -2 & -22 & 3 \\ -45 & 4 & -16 & 6 & -18 & 8 \\ 18 & -26 & -27 & -19 & -48 & -35 \\ 33 & 35 & 9 & 25 & 2 & 7 \\ -8 & 10 & -12 & -34 & 11 & 38 \end{bmatrix}$$

8. SageMathCell 6 Compute $A + Q$

9. SageMathCell 7 Compute $3A + 4T$ [𝕊]-279

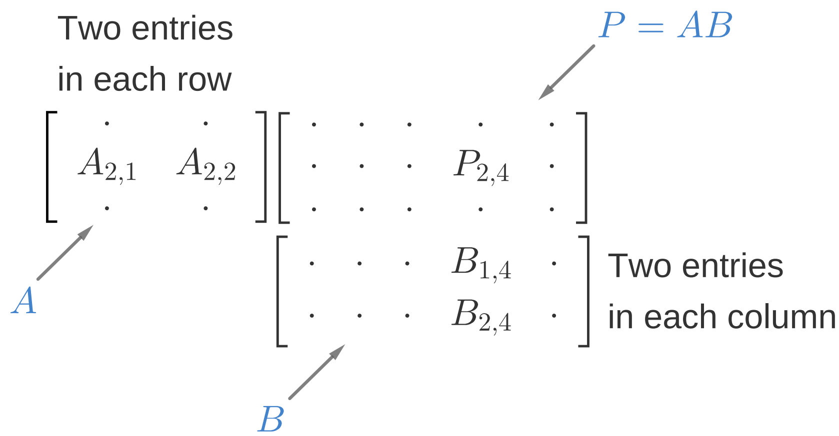
10. SageMathCell 8 Compute $N - 5T$

11. SageMathCell 9 Compute $3.17(1.11Q + .22N)$

Answers

Sudoku sum: Since each block of a sudoku board is required to contain the numbers from 1 through 9 exactly once each, the sum of a single block is $1 + 2 + 3 + \cdots + 9 = \frac{9 \cdot 10}{2} = 45$ making the sum of any pair of blocks 90.

Figure 1.3.1: $(AB)_{2,4} = P_{2,4} = A_{2,1}B_{1,4} + A_{2,2}B_{2,4}$



1.3 Matrix Multiplication

Matrix addition, matrix subtraction and scalar multiplication are each done component-wise, something many people find natural. Even those for whom it does not come naturally rarely question why the operations are done the way they are. After explanation, they are acceptable. Devoid of context, however, there is nothing natural or intuitive about matrix multiplication. It's not difficult. It just takes some getting used to. The purpose of the current section is to start the process of familiarization. The reason multiplication is done the way it is will not come up for a little while yet. In the meantime, a little patience and concentration will be enough.

If you can master the product of a **row matrix** (a $1 \times n$ matrix) with a **column matrix** (an $m \times 1$ matrix), you can master the product of any two matrices. The following example illustrates the process.

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 \end{bmatrix} = \begin{bmatrix} 32 \end{bmatrix} \tag{1.3.1}$$

Given a row matrix R and a column matrix C with the same number of entries, say n, their product is the sum of the products of corresponding entries. That is,

$$RC = \begin{bmatrix} r_{1,1}c_{1,1} + r_{1,2}c_{2,1} + \cdots + r_{1,n}c_{n,1} \end{bmatrix}.$$

The first entry of R (reading from left to right) corresponds with the first entry of C (reading from top to bottom). The second entry of R corresponds with the second entry of C, and so on. The product of the two matrices is the sum of these entry products. As with addition, multiplication is an operator, so the product of two matrices is a matrix. In this case, a 1×1 matrix, as shown in (1.3.1). If R and C differ in length the product RC is undefined.

For matrices with multiple rows and columns, this row-matrix-column-matrix calculation is repeated for each entry of their product. The i, j-entry of $P = AB$ is the single entry of the i^{th} row of A times the j^{th} column of B, where this makes sense. If A and B are matrices, then the product $P = AB$ is calculated by setting $(AB)_{i,j}$ equal to the lone entry of $A_{i,:}B_{:,j}$ (where this makes sense). Several conclusions can be drawn from this description.

- The rows of A and the columns of B must have the same number of entries. Otherwise $A_{i,:}B_{:,j}$, is undefined.
- P has the same number of rows as A (P and A have the same height).
- P has the same number of columns as B (P and B have the same width).

These last two observations suggest an organizational technique for multiplication. Writing B to the right of A and just below leaves a space above B and to the right of A that's exactly the right size for the product P. Plus, the row needed for calculating $(AB)_{i,j}$ is directly left of it and the column needed for calculating $(AB)_{i,j}$ is directly below it. See figure 1.3.1.

Transposition and the Dot Product

If A is a matrix, then its transpose is the matrix resulting from turning the rows of A into columns. The first row of the matrix becomes the first column of the transpose. The second row of the matrix becomes the second column of the transpose, and so on. Equivalently, the transpose of a matrix A is the matrix resulting from turning the columns of A into rows. The first column of the matrix becomes the first row of the transpose. The second column of the matrix becomes the second row of the transpose, and so on. Can you see why turning rows into columns and turning columns into rows are equivalent?

If a matrix has only one row (is a row matrix) then its transpose has one column (is a column matrix), and vice versa. Using a superscript T for transpose the row-matrix-column-matrix product from the beginning of this section can be written

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}^T\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1\cdot 4 + 2\cdot 5 + 3\cdot 6 \end{bmatrix} = \begin{bmatrix} 32 \end{bmatrix} \tag{1.3.2}$$

Writing this way may help you keep track of which numbers should be multiplied by which since they are side by side in the expression using the transpose. Combining this observation with the organizational technique of figure 1.3.1, computing the product

$$\begin{bmatrix} 1 & -2 & 4 \\ 5 & 3 & 6 \end{bmatrix}\begin{bmatrix} -2 & 0 & 9 & 3 \\ 8 & 14 & 2 & 8 \\ 1 & -1 & 7 & 5 \end{bmatrix}$$

might look (at least to start) like the following on paper.

$$\begin{array}{cc} \begin{bmatrix} 1 & -2 & 4 \\ 5 & 3 & 6 \end{bmatrix} & \begin{bmatrix} -14 & -32 & & \\ 20 & & & \end{bmatrix} \\ & \begin{bmatrix} -2 & 0 & 9 & 3 \\ 8 & 14 & 2 & 8 \\ 1 & -1 & 7 & 5 \end{bmatrix} \end{array}$$

$$P_{1,1}: \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}^T\begin{bmatrix} -2 \\ 8 \\ 1 \end{bmatrix} = 1(-2)-2(8)+4(1) = -14$$

$$P_{1,2}: \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}^T\begin{bmatrix} 0 \\ 14 \\ -1 \end{bmatrix} = 1(0)-2(14)+4(-1) = -32$$

$$P_{2,1}: \begin{bmatrix} 5 \\ 3 \\ 6 \end{bmatrix}^T\begin{bmatrix} -2 \\ 8 \\ 1 \end{bmatrix} = 5(-2)+3(8)+6(1) = 20$$

$$P_{2,2}: \begin{bmatrix} 5 \\ 3 \\ 6 \end{bmatrix}^T\begin{bmatrix} 0 \\ 14 \\ -1 \end{bmatrix} =$$

For example, the -32, $P_{1,2}$, is calculated by taking the row directly to its left, $\begin{bmatrix} 1 & -2 & 4 \end{bmatrix}$, and multiplying by the column directly below it, $\begin{bmatrix} 0 \\ 14 \\ -1 \end{bmatrix}$. This product is calculated to the right of the matrices and is just one of the 8 entries of the product. It looks like a lot of work, and it is! Not to worry, though. With some practice, you will become proficient and not have to write down all the individual row-matrix-column-matrix products in such detail. In fact, it will be very important that you acquire such proficiency. This row-matrix-column-matrix calculation sits at the core of linear algebra and its connection to various sciences.

If you have seen the dot product, a very similar calculation in physics or calculus, think of the row-matrix-column-matrix product as the linear algebra equivalent of the dot product.

In physics or calculus (**vectors**): $\langle 5,3,6\rangle \cdot \langle 0,14,-1\rangle = 5\cdot 0 + 3\cdot 14 + 6\cdot -1 = 36$

In linear algebra (matrices): $\begin{bmatrix} 5 \\ 3 \\ 6 \end{bmatrix}^T\begin{bmatrix} 0 \\ 14 \\ -1 \end{bmatrix} = [5\cdot 0 + 3\cdot 14 + 6\cdot -1] = [36]$

It's the same calculation! There are enough similarities between column matrices and vectors that we often use column matrix notation to represent vectors and call them **column vectors** or just vectors, and we call the row-matrix-column-matrix calculation the **dot product**.

Crumpet 7: Row Vector

A row matrix is sometimes referred to as a **row vector** and can be used to represent vectors like those in physics or calculus just as a column vector can.

Thus the distinction between the two objects is blurred, but make no mistake, a column matrix is a matrix, and a vector is a vector. They are not the same thing. It is a convenience in linear algebra to represent vectors as column matrices, giving the column matrix notation two meanings, (1) a matrix, and (2) a vector. Though we try not to do this type of thing often in mathematics, giving a single notation multiple meanings, it happens much like words in English are given multiple meanings. What you can do with a ring depends entirely on what type of ring. A wedding ring might be worn on your ring finger, and a circus ring might contain a tiny car with two dozen clowns in it. Certainly not the other way around!

Crumpet 8: Ring

In mathematics, a ring is a set together with two binary operators that satisfy a number of properties. This is something you will study in abstract algebra.

Analogously, what you can do with a one-column array of numbers depends entirely on what it represents. If it represents a matrix, it might be transposed or used in the solution of a sytem of linear equations. If it represents a vector it might be used in the dot product with another vector or plotted in the Cartesian coordinate system.

Notice the product in equation (1.3.2) is written as a 1×1 matrix, but the same type of matrix product is written as a scalar in the pencil-and-paper calculation of a matrix product. This is another example of a single notation having multiple interpretations, indicated through context. There is no context for equation (1.3.2), so the product is rightfully a matrix. In the calculation of a matrix product, the result of each individual dot product will become an entry—a scalar, not a matrix—in the product. The square brackets are dropped. The 1×1 matrix is treated as if it were a scalar. In fact, 1×1 matrices and scalars are often used interchangeably, jeopardizing the distinction between these two objects. Again, make no mistake, a 1×1 matrix is a matrix, and a scalar is not a matrix at all. They are different things. It is a convenience to let 1×1 matrix notation (square brackets) and scalar notation (lack of delimiters) represent one another, whichever is appropriate for the situation.

Can you compute the products

$$\begin{bmatrix} 1 & -2 \\ 3 & 7 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ -1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 3 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 3 & 7 \end{bmatrix}?$$

Answer on page 19. Besides good practice in multiplying matrices, this example shows that

$$\begin{bmatrix} 1 & -2 \\ 3 & 7 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ -1 & 0 \end{bmatrix} \neq \begin{bmatrix} 2 & 3 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 3 & 7 \end{bmatrix},$$

and more importantly, therefore ***matrix multiplication is not commutative***. Given matrices M and N, we cannot expect MN and NM to be equal even when both products are defined.

Key Concepts

row matrix A matrix with only one row.

column matrix A matrix with only one column.

column vector A vector represented as a column matrix.

row vector A vector represented as a row matrix.

matrix multiplication For any matrices A and B, if the rows of A and the columns of B have the same number of entries, then the product AB is defined. Moreover, AB has the same number of rows (height) as A and the same number of columns (width) as B, and $(AB)_{i,j}$ equals the lone entry of $A_{i,:}B_{:,j}$ for all entries $(AB)_{i,j}$ of AB. If the rows of A and columns of B do not have the same number of entries, then AB is undefined.

transpose For any $m \times n$ matrix A, the transpose of A, denoted A^T, is defined as the $n \times m$ matrix with $(A^T)_{i,j} = a_{j,i}$ for each entry $a_{j,i}$ of A.

vector A quantity with both magnitude and direction.

dot product the dot product of $m \times 1$ matrices $\mathbf{u}$ and $\mathbf{v}$ is $\mathbf{u}^T\mathbf{v}$.

SageMath

If `M` is a matrix in SageMath, then `M.transpose()` is its transpose. The following code defines the matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, extracts columns 2 and 3 as column matrices, and finds the (matrix) product $A_{:,2}^T A_{:,3}$ SageMathCell [10].

```
A=matrix(2,3,[1,2,3,4,5,6])
print("Matrix A:")
print(A)
print()
print("Treating columns 2 and 3 as matrices:")
print()
c2 = matrix(2,1,A.column(1))
c3 = matrix(2,1,A.column(2))
print("column 2:")
print(c2)
print()
print("column 3:")
print(c3)
print()
print("column 2 transpose times column 3:")
print(c2.transpose()*c3)
```

The output of this code is

```
Matrix A:
[1 2 3]
[4 5 6]

Treating columns 2 and 3 as matrices:

column 2:
[2]
[5]

column 3:
[3]
[6]
```

```
column 2 transpose times column 3:
[36]
```

Notice the columns are displayed as column matrices, and the product is also displayed as a matrix, using the square brackets. The `.column()` method extracts a column of a matrix as a vector, however, which is why the definitions of `c2` and `c3` explicitly take each column and feed them to the `matrix()` function.

On the other hand, SageMath is perfectly capable of treating the columns as vectors, as seen in the following code SageMathCell 11. The * operator is used to compute the dot product of two vectors.

```
A=matrix(2,3,[1,2,3,4,5,6])
print("Matrix A:")
print(A)
print()
print("Treating columns 2 and 3 as vectors:")
print()
c2 = A.column(1)
c3 = A.column(2)
print("column 2:")
print(c2)
print()
print("column 3:")
print(c3)
print()
print("Dot product of columns 2 and 3:")
print(c2*c3)
```

The output of this code is

```
Treating columns 2 and 3 as vectors:

column 2:
(2, 5)

column 3:
(3, 6)

Dot product of columns 2 and 3:
36
```

Notice the notation for a vector (parentheses around a comma-separated list of entries), making it clear SageMath is interpreting the columns as vectors, not matrices. Also notice the dot product is displayed (and indeed interpreted) as a scalar, not a matrix.

Exercises

1. Multiply if possible.

(a) $\begin{bmatrix} 3 & 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 5 \\ 7 \end{bmatrix}$

(b) $\begin{bmatrix} 7 & 6 \end{bmatrix}\begin{bmatrix} 4 \\ -5 \end{bmatrix}$

(c) $\begin{bmatrix} -9 \\ -4 \\ 4 \end{bmatrix}\begin{bmatrix} 2 \\ 9 \\ -6 \end{bmatrix}$ [$\mathbb{S}$]-280

(d) $\begin{bmatrix} -1 & 0 & -3 \end{bmatrix}\begin{bmatrix} 6 \\ -2 \\ 5 \end{bmatrix}$ [$\mathbb{S}$]-280

(e) $\begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix}\begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}$

(f) $\begin{bmatrix} -3 & 2 \end{bmatrix}\begin{bmatrix} 7 \\ -1 \end{bmatrix}$ [$\mathbb{A}$]-347

(g) $\begin{bmatrix} 5.8 & 0.2 \end{bmatrix}\begin{bmatrix} 2.5 \\ 3.8 \end{bmatrix}$ [$\mathbb{A}$]-347

(h) $\begin{bmatrix} -3 & 3 & 4 \end{bmatrix}\begin{bmatrix} -1 \\ 5 \end{bmatrix}$

(i) $\begin{bmatrix} 7 & 8 & 2 & 4 \end{bmatrix}\begin{bmatrix} -6 & 1 & 6 & 4 \end{bmatrix}$

(j) $\begin{bmatrix} 7 & 0 & 4 & 6 \end{bmatrix}\begin{bmatrix} -2 \\ 7 \\ 1 \\ 3 \end{bmatrix}$

(k) $\begin{bmatrix} 1.35 & 4.58 & 7.36 \end{bmatrix}\begin{bmatrix} 3.36 & -0.25 & 1.6 \end{bmatrix}$

2. Multiply if possible.

(a) $\begin{bmatrix} 3 & 0 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} 2 & 7 \\ 6 & 3 \end{bmatrix}$

(b) $\begin{bmatrix} 7 \\ -2 \end{bmatrix}\begin{bmatrix} -2 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 0.03 & -0.6 \\ 4.25 & 5.09 \end{bmatrix}\begin{bmatrix} -0.3 \\ 4.6 \end{bmatrix}$

(d) $\begin{bmatrix} 6.3 \\ 4.1 \\ 3.4 \end{bmatrix}\begin{bmatrix} 2.3 & 4.5 \end{bmatrix}$ [𝕊]-280

(e) $\begin{bmatrix} 1 & 9 & 10 \\ 3 & 0 & 8 \\ 3 & 8 & 10 \end{bmatrix}\begin{bmatrix} 8 & 4 \\ 10 & 10 \\ 9 & 5 \\ 3 & 8 \end{bmatrix}$ [𝕊]-280

(f) $\begin{bmatrix} -3 & 0 & 1 \\ 2 & 5 & 7 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$ [𝕊]-280

(g) $\begin{bmatrix} 2 & 3 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix}$ [𝔸]-347

(h) $\begin{bmatrix} 6 & 7 & 4 \\ -3 & 0 & 7 \end{bmatrix}\begin{bmatrix} 3 & 4 \\ -2 & 1 \\ -1 & 2 \end{bmatrix}$

(i) $\begin{bmatrix} 4 \\ -1 \end{bmatrix}\begin{bmatrix} 5 & 1 & 4 & 6 \end{bmatrix}$

(j) $\begin{bmatrix} 7.94 \\ 1.15 \\ 2.88 \\ 8.95 \end{bmatrix}\begin{bmatrix} 9.98 & 2.91 \\ 1.48 & 8.05 \\ 6.41 & 9.67 \\ 5.16 & 8.88 \end{bmatrix}$

(k) $\begin{bmatrix} 3 & 0 \\ 5 & -4 \end{bmatrix}\begin{bmatrix} 6 & 3 & 1 \\ 6 & 1 & 7 \end{bmatrix}$ [𝔸]-347

(l) $\begin{bmatrix} 2 & 5 \\ -1 & 0 \\ -3 & 4 \end{bmatrix}\begin{bmatrix} 4 & 2 \\ 6 & -3 \end{bmatrix}$

(m) $\begin{bmatrix} 8 \\ 0 \\ 9 \end{bmatrix}\begin{bmatrix} 2 & 4 & -1 & 1 \\ 9 & 6 & 10 & 8 \end{bmatrix}$ [𝔸]-347

(n) $\begin{bmatrix} 0 & 0 & 3 & 6 \\ -3 & 0 & 7 & -3 \\ 1 & -1 & 4 & 4 \end{bmatrix}\begin{bmatrix} 4 \\ 7 \\ 4 \\ 5 \end{bmatrix}$

(o) $\begin{bmatrix} 5 & 4 \end{bmatrix}\begin{bmatrix} -1 & 2 \\ 6 & -1 \end{bmatrix}$

(p) $\begin{bmatrix} 3.47 & -2.73 \end{bmatrix}\begin{bmatrix} 5.53 & 5.89 \\ 5.24 & 0.82 \end{bmatrix}$

(q) $\begin{bmatrix} 1 & -1 & 6 & 3 \\ 10 & 4 & 8 & 3 \end{bmatrix}\begin{bmatrix} 8 & 3 & 7 & 10 \\ 8 & 1 & 4 & 4 \end{bmatrix}$

3. Find the dot product $\mathbf{u}^T\mathbf{v}$.

(a) $\mathbf{u} = \begin{bmatrix} -7 \\ 8 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 9 \\ -3 \end{bmatrix}$

(b) $\mathbf{u} = \begin{bmatrix} -11 \\ 3 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 11 \\ 13 \end{bmatrix}$

(c) $\mathbf{u} = \begin{bmatrix} -10 \\ 3 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 0 \\ 9 \end{bmatrix}$ [𝔸]-347

(d) $\mathbf{u} = \begin{bmatrix} 14.3 \\ -13.7 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 10.3 \\ 2.9 \end{bmatrix}$

(e) $\mathbf{u} = \begin{bmatrix} 10 \\ 2 \\ -3 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -11 \\ 3 \\ -10 \end{bmatrix}$ [𝕊]-281

(f) $\mathbf{u} = \begin{bmatrix} 2 \\ -6 \\ 12 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -6 \\ -10 \\ -4 \end{bmatrix}$

(g) $\mathbf{u} = \begin{bmatrix} 8 \\ -7 \\ 5 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 5 \\ -11 \\ -8 \end{bmatrix}$ [𝔸]-347

(h) $\mathbf{u} = \begin{bmatrix} 4.9 \\ 0.4 \\ -2.5 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 3.6 \\ 2.0 \\ -4.1 \end{bmatrix}$

(i) $\mathbf{u} = \begin{bmatrix} -1 \\ 7 \\ 0 \\ 2 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -7 \\ -5 \\ -4 \\ -2 \end{bmatrix}$

(j) $\mathbf{u} = \begin{bmatrix} 3 \\ -3 \\ 7 \\ -2 \\ -8 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 0 \\ -2 \\ -1 \\ 5 \\ 4 \end{bmatrix}$ [𝔸]-347

4. Redo question 3 calculating $\mathbf{v}^T\mathbf{u}$ instead, and compare your answers. [𝕊]-281 [𝔸]-347

5. Suppose A is a matrix of size 2×7, C is a matrix of size 5×7, and the matrix computation $A + BC$ is defined. What is the size of matrix B? [𝔸]-347

6. Matrices A, B, C, D are such that $(A + B)(CD)$ is defined (all of the operations are possible). If B is a 3×4 matrix and D is a 5×8 matrix, what are the dimensions of A and C?

7. Describe how to multiply two matrices, and explain how to determine whether the multiplication can be done.

8. True or false? For any column vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ with the same number of entries in each, [𝔸]-347

(a) $\mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u}$

(b) $(\mathbf{u} + \mathbf{w})^T\mathbf{v} = \mathbf{u}^T\mathbf{v} + \mathbf{w}^T\mathbf{v}$

9. Find a pair of matrices M and N so that MN is defined, but NM is not, and therefore $MN \neq NM$.

10. Find a pair of matrices M and N such that MN and NM are both defined but are different sizes, and therefore $MN \neq NM$.

11. Find a pair of 3×3 matrices M and N such that $MN \neq NM$.

12. Can you find a pair of distinct 2×2 matrices M and N such that $MN = NM$?

13. Suppose the matrix product MN is defined (the multiplication can be done). Which of the following is true?

    (a) M and N must have the same number of rows.

    (b) M and N must have the same number of columns.

    (c) The number of rows of M must equal the number of columns of N .

    (d) The number of columns of M must equal the number of rows of N .

    (e) None of the above.

14. Find the dot product, $\mathbf{u}^T\mathbf{v}$.

    (a) SageMathCell 12 [$\mathbb{S}$]-281

    $$\mathbf{u} = \begin{bmatrix} -3748 \\ -3468 \\ -4357 \\ -3611 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -5497 \\ 2448 \\ -2990 \\ -2772 \end{bmatrix}$$

    (b) SageMathCell 13

    $$\mathbf{u} = \begin{bmatrix} -1.33017 \\ 1.33699 \\ 5.50693 \\ 9.67517 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 9.21163 \\ 2.87319 \\ -9.634 \\ 4.46961 \end{bmatrix}$$

    (c) SageMathCell 14

    $$\mathbf{u} = \begin{bmatrix} -228 \\ -5201 \\ -451 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -8419 \\ -5162 \\ -2381 \end{bmatrix}$$

    (d) SageMathCell 15

    $$\mathbf{u} = \begin{bmatrix} -2.6018 \\ 5.18949 \\ 2.99411 \\ 7.25436 \\ -0.90284 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -7.29805 \\ 1.89209 \\ 7.33303 \\ -9.41897 \\ 0.85775 \end{bmatrix}$$

For the remaining exercises, let

$$A = \begin{bmatrix} 42 & 0 & -47 & -34 & -10 & -48 \\ 8 & 26 & 43 & -18 & -20 & -30 \\ -41 & -40 & -29 & -36 & -44 & 12 \\ -42 & 47 & 28 & 4 & 38 & -22 \\ 18 & -15 & -1 & 29 & 37 & 9 \end{bmatrix} \quad U = \begin{bmatrix} -21 & -33 & 28 & -15 & 34 & 45 \\ 27 & 40 & -13 & -23 & -10 & 15 \\ 43 & -6 & 46 & 17 & 13 & 21 \\ -40 & -46 & 2 & 16 & 22 & -14 \\ 10 & -12 & 29 & 35 & 48 & -31 \end{bmatrix}$$

$$Q = \begin{bmatrix} -17 & -37 & -34 & 20 & -14 & 10 \\ -23 & 44 & 47 & 18 & 19 & 49 \\ 11 & 33 & 35 & -50 & 2 & 9 \\ -36 & -18 & 7 & 17 & -49 & 31 \\ -8 & 16 & 28 & -32 & -2 & 5 \end{bmatrix} \quad R = \begin{bmatrix} 40 & 47 & 13 & -2 & -22 & 3 \\ -45 & 4 & -16 & 6 & -18 & 8 \\ 18 & -26 & -27 & -19 & -48 & -35 \\ 33 & 35 & 9 & 25 & 2 & 7 \\ -8 & 10 & -12 & -34 & 11 & 38 \end{bmatrix}$$

14. SageMathCell 16 Compute $(A^T)(U)$ and $(U)(A^T)$. Are they equal?

15. SageMathCell 17 Compute Q^TR and QR^T. Are they equal? [$\mathbb{S}$]-281

16. SageMathCell 18 Compute $\left(3Q^T - 2R^T\right)^T$ and $3Q - 2R$. What do you notice? Why?

17. SageMathCell 19 Can you determine which of the following computations are defined? Ask SageMath to compute them all. The ones that are undefined will produce long error messages.

$$(QR)^T \qquad AU^TR \qquad QUAR^T \qquad A^TQU^TA$$
$$A + R^T \qquad AR^T \qquad (R - A)^T$$

Answers

matrix products The products are

$$\begin{bmatrix} 1 & -2 \\ 3 & 7 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ -1 & 9 \end{bmatrix}$$
$$\begin{bmatrix} 2 & 3 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 3 & 7 \end{bmatrix} = \begin{bmatrix} 11 & 17 \\ -1 & 2 \end{bmatrix}.$$

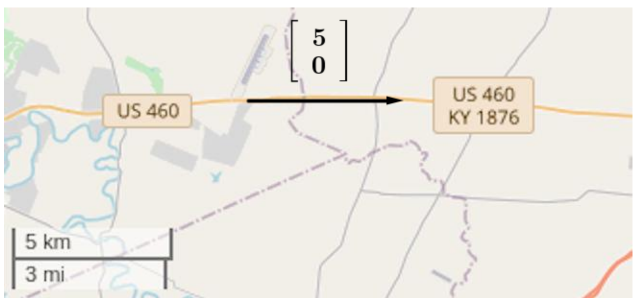
1.4 Magnitude and Orthogonality

Geometric Interpretation of Vectors

One day my friend Victor took a 5 kilometer drive. When Victor told me this I knew just how long his drive was. It was 5 kilometers. When Victor added that his drive was on a very straight highway headed due east, I knew more. I knew which way Victor was driving. I could imagine tracing out his path on a map by drawing a horizontal arrow pointing to the right (eastward) with a magnitude equivalent to 5 kilometers. The arrow captures both the direction and magnitude of Victor's drive. Vectors can be imagined in the same way. The vector

$$\begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

has a 5 as its first entry and a 0 as its second. Thinking of these entries as x- and y- coordinates, the five represents 5 units right (eastward) and the zero represents 0 units up (northward). In this way, the vector represents both the magnitude and direction of Victor's drive, just like the arrow. The vector and the arrow can be interpreted to represent the same thing, blurring any distinction between them. [3]

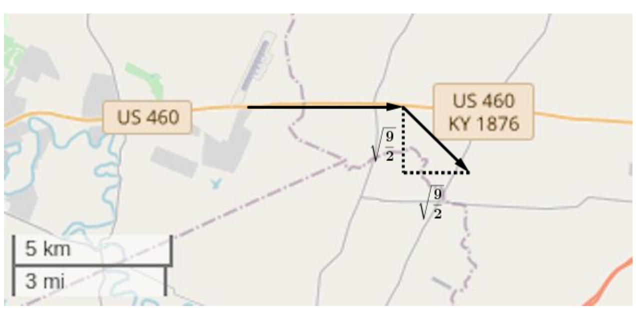


The vector/arrow represents Victor's displacement, or movement, 5 kilometers in the eastward direction.

Notice there is no origin on the map. This is typical of drawing vectors. They are not specified relative to an origin. They only represent a change in location, or displacement, starting anywhere. A vector represents the locations of two points relative to one another. Exactly where those two points lie is not determined by the vector itself. Further information is needed to locate the vector. In the case of Victor's travel, I needed to know on what road and where he was driving to create an accurate picture of his drive.

After driving 5 kilometers east, Victor exited the highway and drove 3 kilometers southeast (using a road that does not appear on the map). When I heard this, I was able to capture this part of Victor's journey by the vector

$$\begin{bmatrix} \sqrt{\frac{9}{2}} \\ -\sqrt{\frac{9}{2}} \end{bmatrix}.$$



And I knew exactly where to put it since it started just where the previous leg left off. Drawn as an arrow, the vector is the hypotenuse of a right triangle with side lengths $\sqrt{\frac{9}{2}}$, which by the Pythagorean theorem gives it length

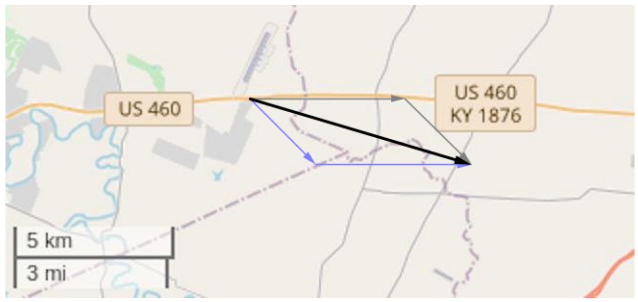
[3] Street map minus vectors © OpenStreetMap contributors

$\sqrt{\left(\sqrt{\frac{9}{2}}\right)^2 + \left(\sqrt{\frac{9}{2}}\right)^2} = 3$. It has the right magnitude and it points southeast (and starts where the first leg leaves off), so it accurately represents the second leg of Victor's drive.

As the crow flies, Victor's total displacement or movement for the drive is represented by the sum of the vectors,

$$\begin{bmatrix} 5 \\ 0 \end{bmatrix} + \begin{bmatrix} \sqrt{\frac{9}{2}} \\ -\sqrt{\frac{9}{2}} \end{bmatrix} = \begin{bmatrix} 5 + \sqrt{\frac{9}{2}} \\ -\sqrt{\frac{9}{2}} \end{bmatrix},$$

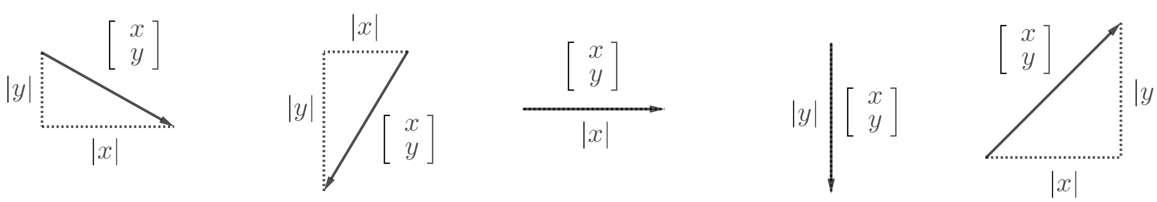
the black vector in the diagram.



Since addition of vectors is commutative, it does not matter which vector is plotted first. In the diagram, the gray vectors represent $\begin{bmatrix} 5 \\ 0 \end{bmatrix} + \begin{bmatrix} \sqrt{\frac{9}{2}} \\ -\sqrt{\frac{9}{2}} \end{bmatrix}$ and the blue vectors represent $\begin{bmatrix} \sqrt{\frac{9}{2}} \\ -\sqrt{\frac{9}{2}} \end{bmatrix} + \begin{bmatrix} 5 \\ 0 \end{bmatrix}$. After the pair of displacements, $\begin{bmatrix} 5 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} \sqrt{\frac{9}{2}} \\ -\sqrt{\frac{9}{2}} \end{bmatrix}$, Victor's total displacement is $\begin{bmatrix} 5 + \sqrt{\frac{9}{2}} \\ -\sqrt{\frac{9}{2}} \end{bmatrix}$ no matter which displacement comes first. The diagram illustrates the parallelogram rule for vector addition. The sum of two vectors is a diagonal of the parallelogram determined by the two vectors.

Perpendicularity

The magnitude of a vector, not surprisingly, is defined by the length of its representative arrow. A collection of vectors pointing in various directions, including vertical and horizontal are shown below.



Regardless of which direction the vector $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ points, its magnitude is $\sqrt{|x|^2 + |y|^2}$ or simply $\sqrt{x^2 + y^2}$. The Pythagorean theorem can be used to calculate magnitudes of vectors that are not horizontal or vertical.

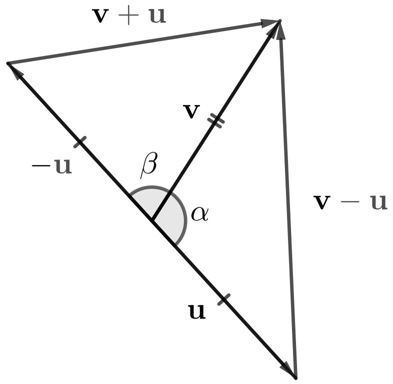
Coincidentally the dot product of $\mathbf{v}$ with itself, $\mathbf{v}^T\mathbf{v}$, is

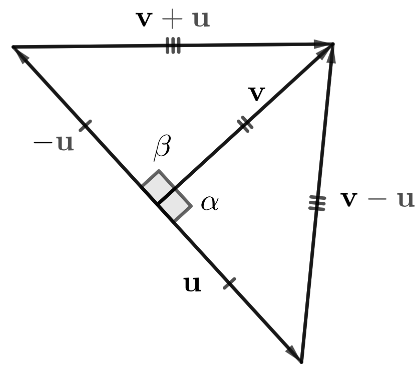
$$\begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + y^2$$

so the magnitude of $\mathbf{v}$ can also be written as $\sqrt{\mathbf{v}^T\mathbf{v}}$. This expression has a nice symmetry and is independent of the number of entries in $\mathbf{v}$. It could apply to vectors with 3, 8, or 28 entries just as well as vectors with 2 entries. The **magnitude** of a vector $\mathbf{v}$, denoted $||\mathbf{v}||$, is defined as

$$||\mathbf{v}|| = \sqrt{\mathbf{v}^T\mathbf{v}}.$$

The following diagram illustrates the relationship between the magnitudes of vectors $\mathbf{v} - \mathbf{u}$ and $\mathbf{v} + \mathbf{u}$. By the side-angle-side theorem from geometry the pair of triangles in each figure are congruent if and only if $\alpha = \beta$. Since α and β together form a straight angle, $\alpha = \beta$ if and only if they are both right angles. Consequently the magnitudes of $\mathbf{v} + \mathbf{u}$ and $\mathbf{v} - \mathbf{u}$ are equal if and only if $\mathbf{u}$ and $\mathbf{v}$ are perpendicular.





This observation leads to a very useful property of the dot product, exposed by the following calculation. $\mathbf{u}$ and $\mathbf{v}$ are perpendicular if and only if

$$\begin{aligned}
\|\mathbf{v}+\mathbf{u}\| &= \|\mathbf{v}-\mathbf{u}\| \\
\sqrt{(\mathbf{v}+\mathbf{u})^T(\mathbf{v}+\mathbf{u})} &= \sqrt{(\mathbf{v}-\mathbf{u})^T(\mathbf{v}-\mathbf{u})} \\
(\mathbf{v}+\mathbf{u})^T(\mathbf{v}+\mathbf{u}) &= (\mathbf{v}-\mathbf{u})^T(\mathbf{v}-\mathbf{u}) \\
(\mathbf{v}^T+\mathbf{u}^T)(\mathbf{v}+\mathbf{u}) &= (\mathbf{v}^T-\mathbf{u}^T)(\mathbf{v}-\mathbf{u}) \\
\mathbf{v}^T\mathbf{v}+\mathbf{v}^T\mathbf{u}+\mathbf{u}^T\mathbf{v}+\mathbf{u}^T\mathbf{u} &= \mathbf{v}^T\mathbf{v}-\mathbf{v}^T\mathbf{u}-\mathbf{u}^T\mathbf{v}+\mathbf{u}^T\mathbf{u} \\
\mathbf{v}^T\mathbf{u}+\mathbf{u}^T\mathbf{v} &= -\mathbf{v}^T\mathbf{u}-\mathbf{u}^T\mathbf{v} \\
2\mathbf{v}^T\mathbf{u} &= -2\mathbf{u}^T\mathbf{v} \\
2\mathbf{v}^T\mathbf{u} &= -2\mathbf{v}^T\mathbf{u} \\
4\mathbf{v}^T\mathbf{u} &= 0 \\
\mathbf{v}^T\mathbf{u} &= 0
\end{aligned} \tag{1.4.1}$$

Since each line follows logically from the previous, and vice veresa, the vectors $\mathbf{u}$ and $\mathbf{v}$ (with two entries) are perpendicular if and only if their dot product is zero! Passing between the seventh equation and the eighth depends on the fact that $\mathbf{v}^T\mathbf{u} = \mathbf{u}^T\mathbf{v}$. Can you show this is true for any vectors of equal size? Answer on page 26.

As with the formula $\|\mathbf{v}\| = \sqrt{\mathbf{v}^T\mathbf{v}}$ for magnitude, this calculation is independent of the number of entries in the vectors. We say that vectors $\mathbf{u}$ and $\mathbf{v}$ of the same size are **orthogonal** if and only if their dot product is zero. For vectors with two or three entries this means the vectors are perpendicular. As a result, orthogonality is precisely the same as perpendicularity in two and three dimensions, and extends the idea to dimensions greater than three.

If $\mathbf{u}$ and $\mathbf{v}$ are placed with their tails at the same point, then $\|\mathbf{u} - \mathbf{v}\|$ is the distance between the heads of $\mathbf{u}$ and $\mathbf{v}$. See the diagram above. As such, the distance between $\mathbf{u}$ and $\mathbf{v}$, denoted $d(\mathbf{u}, \mathbf{v})$, is defined as $\|\mathbf{u} - \mathbf{v}\|$. Easy to picture in two dimensions, this formula applies to vectors of any magnitude again extending a two- and three-dimensional notion to higher dimensions.

Key Concepts

geometric interpretation of vectors Vectors are often thought of as displacements represented by arrows.

geometric interpretation of vector sum The sum of two vectors is represented geometrically by a diagonal of the parallelogram determined by the two vectors.

magnitude of a column vector $\mathbf{v}$, denoted $\|\mathbf{v}\|$, is the square root of the dot product of $\mathbf{v}$ with itself, $\sqrt{\mathbf{v}^T\mathbf{v}}$.

orthogonal Two vectors whose dot product is defined and zero are orthogonal.

distance The distance between two vectors, $d(\mathbf{u}, \mathbf{v})$, is the magnitude of their difference, $\|\mathbf{u} - \mathbf{v}\|$.

SageMath

SageMath distinguishes between vectors and matrices, but just like in mathematics the distinction is blurry. The SageMath code

```
u=vector([1,2,3])
v=matrix(3,1,[3,2,1])
print(u*v)
```

SageMathCell 20 runs even though the third line requests the product of a vector with a matrix. SageMath treats matrix `v` as if it were a vector, sort of. The output of the code is

```
(10)
```

a vector with one entry—-not a scalar and not a 1×1 matrix. If `v` is defined as a vector as in the following code, the output is the scalar value 10, not a vector.

```
u=vector([1,2,3])
v=vector([3,2,1])
print(u*v)
```

SageMathCell 21 produces

```
10
```

SageMath's internal process of converting one type of variable to another to avoid throwing an error, a process called coersion, can produce unanticipated results. More predictable results are obtained by explicitly converting one type of variable to another. The SageMath code

```
u=vector([1,2,3])
v=matrix(3,1,[3,2,1])
print(u*vector(v))
```

SageMathCell 22 explicitly tells SageMath to treat `v` as a vector in the computation of the product so no coersion is needed, and it produces

```
10
```

just as if `v` were defined as a vector in the first place.

Any row or column matrix can be converted to a vector the same way. In fact, vectors can be converted to row or column matrices just as easily. The following code converts `u` to a matrix (instead of converting `v` to a vector) and then computes the dot product.

```
u=vector([1,2,3])
v=matrix(3,1,[3,2,1])
print(matrix(1,3,u)*v)
```

SageMathCell 23 produces the 1×1 matrix

```
[10]
```

since the multiplicands are both matrices. Be aware that vectors and matrices are not equivalent in SageMath. Unexpected results may be seen when the two types are intermingled. To avoid surprises, convert one to the other explicitly as needed.

The magnitude of a vector can be computed using the `.norm()` method. Consistent with the developing theme, the `.norm()` method can be applied to either matrices or vectors, and the results are different! The following code defines the "same" vector as both a SageMath vector and a SageMath matrix and then outputs their magnitudes, or norms.

```
u_vec=vector([6,5,-3])
u_mat=matrix(1,3,[6,5,-3])
print(u_vec.norm())
print(u_mat.norm())
```

SageMath Cell 24 produces

```
sqrt(70)
8.366600265340756
```

The norm of a vector is computed symbolically while the norm of a matrix is computed as an approximate decimal equivalent. $\sqrt{70} \approx 8.366600265340756$.

Exercises

1. Calculate $\|\mathbf{u}\|$.

   (a) $\mathbf{u} = \begin{bmatrix} -7 \\ 8 \end{bmatrix}$

   (b) $\mathbf{u} = \begin{bmatrix} -11 \\ 3 \end{bmatrix}$

   (c) $\mathbf{u} = \begin{bmatrix} -10 \\ 3 \end{bmatrix}$ [𝔸]-347

   (d) $\mathbf{u} = \begin{bmatrix} 14.3 \\ -13.7 \end{bmatrix}$

   (e) $\mathbf{u} = \begin{bmatrix} 10 \\ 2 \\ -3 \end{bmatrix}$

   (f) $\mathbf{u} = \begin{bmatrix} 2 \\ -6 \\ 12 \end{bmatrix}$ [𝕊]-282

   (g) $\mathbf{u} = \begin{bmatrix} 8 \\ -7 \\ 5 \end{bmatrix}$

   (h) $\mathbf{u} = \begin{bmatrix} 4.9 \\ 0.4 \\ -2.5 \end{bmatrix}$ [𝔸]-347

   (i) $\mathbf{u} = \begin{bmatrix} -1 \\ 7 \\ 0 \\ 2 \end{bmatrix}$

   (j) $\mathbf{u} = \begin{bmatrix} 3 \\ -3 \\ 7 \\ -2 \\ -8 \end{bmatrix}$ [𝔸]-347

2. Calculate $d(\mathbf{u}, \mathbf{v})$.

   (a) $\mathbf{u} = \begin{bmatrix} -7 \\ 8 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -9 \\ 3 \end{bmatrix}$

   (b) $\mathbf{u} = \begin{bmatrix} -11 \\ 3 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -11 \\ 13 \end{bmatrix}$

   (c) $\mathbf{u} = \begin{bmatrix} -10 \\ 3 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 0 \\ 9 \end{bmatrix}$ [𝔸]-347

   (d) $\mathbf{u} = \begin{bmatrix} 14.3 \\ -13.7 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 10.3 \\ 2.9 \end{bmatrix}$

   (e) $\mathbf{u} = \begin{bmatrix} 10 \\ 2 \\ -3 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -11 \\ 3 \\ -10 \end{bmatrix}$

   (f) $\mathbf{u} = \begin{bmatrix} 2 \\ -6 \\ 12 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -6 \\ -10 \\ 4 \end{bmatrix}$ [𝕊]-282

   (g) $\mathbf{u} = \begin{bmatrix} 8 \\ -7 \\ 5 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 5 \\ -11 \\ -2 \end{bmatrix}$

   (h) $\mathbf{u} = \begin{bmatrix} 4.9 \\ 0.4 \\ -2.5 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 3.6 \\ 2.0 \\ -4.1 \end{bmatrix}$ [𝔸]-347

   (i) $\mathbf{u} = \begin{bmatrix} -1 \\ 7 \\ 0 \\ 2 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -7 \\ -5 \\ -4 \\ 2 \end{bmatrix}$

   (j) $\mathbf{u} = \begin{bmatrix} 3 \\ -3 \\ 7 \\ -2 \\ -8 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 0 \\ -2 \\ 3 \\ 5 \\ -4 \end{bmatrix}$ [𝔸]-347

3. Are $\mathbf{u}$ and $\mathbf{v}$ orthogonal?

   (a) $\mathbf{u} = \begin{bmatrix} 5 \\ 8 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$

   (b) $\mathbf{u} = \begin{bmatrix} -11 \\ 7 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 11 \\ 13 \end{bmatrix}$

(c) $\mathbf{u} = \begin{bmatrix} -10 \\ 8 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$ [$\mathbb{A}$]-347

(d) $\mathbf{u} = \begin{bmatrix} 14.5 \\ -17 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 10.2 \\ 8.7 \end{bmatrix}$

(e) $\mathbf{u} = \begin{bmatrix} 10 \\ 8 \\ -9 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -11 \\ 3 \\ -9 \end{bmatrix}$

(f) $\mathbf{u} = \begin{bmatrix} 2 \\ -6 \\ 12 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -6 \\ -10 \\ -4 \end{bmatrix}$ [$\mathbb{S}$]-282

(g) $\mathbf{u} = \begin{bmatrix} 8 \\ -7 \\ -11 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 5 \\ 12 \\ -4 \end{bmatrix}$

(h) $\mathbf{u} = \begin{bmatrix} 4.9 \\ 0.4 \\ -2.5 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 3.6 \\ 2.0 \\ -4.1 \end{bmatrix}$ [$\mathbb{A}$]-347

(i) $\mathbf{u} = \begin{bmatrix} -1 \\ 7 \\ 0 \\ 2 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} -7 \\ -5 \\ -4 \\ -2 \end{bmatrix}$

(j) $\mathbf{u} = \begin{bmatrix} 3 \\ -3 \\ 7 \\ -3 \\ -8 \end{bmatrix}; \mathbf{v} = \begin{bmatrix} 0 \\ -2 \\ -1 \\ 5 \\ -2 \end{bmatrix}$ [$\mathbb{A}$]-347

4. Find k so that the vectors are orthogonal.

   (a) $\begin{bmatrix} -3 \\ 6 \end{bmatrix}$ and $\begin{bmatrix} -12 \\ k \end{bmatrix}$

   (b) $\begin{bmatrix} 15 \\ 9 \end{bmatrix}$ and $\begin{bmatrix} k \\ 7 \end{bmatrix}$

   (c) $\begin{bmatrix} -2 \\ -6 \\ -3 \end{bmatrix}$ and $\begin{bmatrix} -7 \\ k \\ -10 \end{bmatrix}$ [$\mathbb{S}$]-282

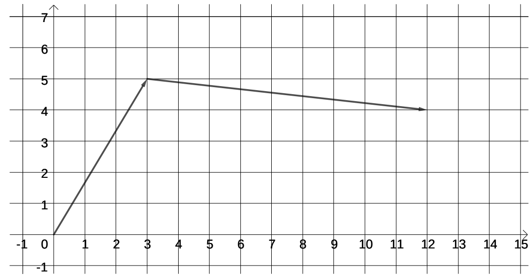
   (d) $\begin{bmatrix} 7 \\ -10 \\ k \end{bmatrix}$ and $\begin{bmatrix} k \\ -4 \\ 6 \end{bmatrix}$

   (e) $\begin{bmatrix} k \\ k \\ 11 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} -7 \\ k \\ 3 \\ -7 \end{bmatrix}$

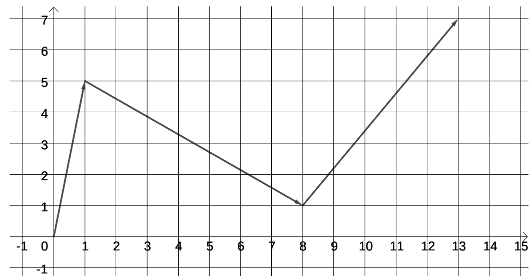
   (f) $\begin{bmatrix} k \\ 2 \\ -14 \\ 4 \\ 6 \\ 10 \\ -10 \\ 8 \\ k \end{bmatrix}$ and $\begin{bmatrix} k \\ 12 \\ 13 \\ 9 \\ -1 \\ 7 \\ 5 \\ 3 \\ 8 \end{bmatrix}$ [$\mathbb{A}$]-347

5. Find the sum of the vectors.

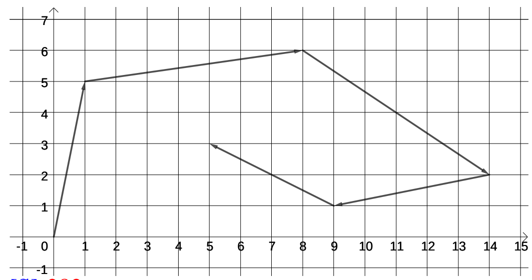
   (a)



   (b)



   (c)



[$\mathbb{S}$]-282

6. Let $\mathbf{u} = \begin{bmatrix} -6 \\ -12 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 9 \\ -5 \end{bmatrix}$. Find a column vector $\mathbf{w}$ such that

   (a) $d(\mathbf{u}+\mathbf{v}, \mathbf{w}) = 1$

   (b) $d(\mathbf{u}+\mathbf{v}, \mathbf{w}) = \sqrt{2}$

   (c) $d(\mathbf{u}+\mathbf{w}, \mathbf{v}) = 1$ [$\mathbb{S}$]-282

   (d) $d(\mathbf{v}+\mathbf{w}, \mathbf{u}) = 1$

   HINT: Make a sketch.

7. What conditions on a column vector $\mathbf{u}$ will make $\mathbf{u}^T\mathbf{u}$ zero?

8. Give an example of a 5×1 column vector $\mathbf{u}$ and 2×1 column vector $\mathbf{v}$ such that the magnitude of $\mathbf{u}$ is less than the magnitude of $\mathbf{v}$.

9. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ and set

$$\mathbf{w} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix}.$$

   (a) Calculate $\mathbf{u}^T\mathbf{w}$.

   (b) Calculate $\mathbf{v}^T\mathbf{w}$.

   (c) Are $\mathbf{u}$ and $\mathbf{w}$ perpendicular?

   (d) Are $\mathbf{v}$ and $\mathbf{w}$ perpendicular?

10. Suppose **u** and **v** are orthogonal.
    (a) Are $3\mathbf{u}$ and $4\mathbf{v}$ orthogonal?
    (b) Are $-12.1\mathbf{u}$ and $0.12\mathbf{v}$ orthogonal? [𝕊]-283

11. SageMathCell 25 Add code that will calculate the norm of the third column of D (treated as a vector). [𝕊]-283

```
D = matrix(3,4,[-2,10,-7,8,-2,7,
                11,-7,-4,5,6,10])
```

What is the output of your code?

12. SageMathCell 26 Add code that will calculate

```
u = vector([3.9,7.2,-8.4,-11.8,.5,
            -11.0,-9.5,8.6])
v = vector([-10.0,10.7,7.1,-6.6,
            11.6,-1.0,12.2,-13.3])
```

(a) $\|\mathbf{u}\|$

(b) $d(\mathbf{u}, \mathbf{v})$

(c) $\mathbf{u} + \mathbf{v}$

What is the output of your code?

Answers

dot product equality Letting $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$

$$\mathbf{u}^T\mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}^T \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

and

$$\mathbf{v}^T\mathbf{u} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}^T \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = v_1u_1 + v_2u_2 + \cdots + v_nu_n.$$

Since multiplication of scalars is commutative, these expressions are equal.

1.5 The Determinant

$b^2 - 4ac$ is "the discriminant", but why? Each quadratic function, $p(x) = ax^2 + bx + c$, has two real roots, one (repeated) real root, or two complex roots. The discriminant discriminates between which quadratics are which. If the coefficients a, b, c, of a quadratic function are such that $b^2 - 4ac > 0$, then the quadratic has two real roots (and no others). If the coefficients are such that $b^2 - 4ac = 0$, then the quadratic has one real root (and no others). If the coefficients are such that $b^2 - 4ac < 0$, then the quadratic has two complex roots (and no others). In this way, the quantity $b^2 - 4ac$ associated with the quadratic function $p(x) = ax^2 + bx + c$ determines what type of roots p has. It is determinative of the type of roots, and in this light might just as well be known as a determinant (which means determinative). In mathematics, though, the term *determinant* is reserved for linear algebra. The determinant is a determinative calculation that can be made for any matrix much the same way the discriminant is a determinative calculation that can be made for any quadratic function. Exactly what the determinant determines will have to wait a short while.

The determinant of an $m \times n$ matrix is undefined if $m \neq n$, so determinants are calculated only for **square matrices**, those with the same number of columns as rows. The determinant of a 1×1 matrix is its lone entry. That is, the determinant of $\begin{bmatrix} a \end{bmatrix}$ is a. As such, the determinant is a scalar. The notations $\det A$ or $|A|$ are used to denote the determinant of the matrix A.

The determinant of a square matrix with more than one row, and therefore more than one column, is defined recursively. If A has n rows and n columns, $n > 1$, then[4]

$$\det A = (-1)^{1+1} A_{1,1} \det A_{\setminus 1,1} + (-1)^{1+2} A_{1,2} \det A_{\setminus 1,2} + \cdots + (-1)^{1+n} A_{1,n} \det A_{\setminus 1,n}. \tag{1.5.1}$$

For example, if $A = \begin{bmatrix} -12 & 49 & -45 & -10 \\ 28 & 45 & -46 & 23 \\ -15 & -28 & 4 & -48 \\ -1 & 34 & -38 & -18 \end{bmatrix}$, then

$$\begin{aligned} \det A = \begin{vmatrix} -12 & 49 & -45 & -10 \\ 28 & 45 & -46 & 23 \\ -15 & -28 & 4 & -48 \\ -1 & 34 & -38 & -18 \end{vmatrix} &= -12 \begin{vmatrix} 45 & -46 & 23 \\ -28 & 4 & -48 \\ 34 & -38 & -18 \end{vmatrix} - 49 \begin{vmatrix} 28 & -46 & 23 \\ -15 & 4 & -48 \\ -1 & -38 & -18 \end{vmatrix} \\ &\quad - 45 \begin{vmatrix} 28 & 45 & 23 \\ -15 & -28 & -48 \\ -1 & 34 & -18 \end{vmatrix} + 10 \begin{vmatrix} 28 & 45 & -46 \\ -15 & -28 & 4 \\ -1 & 34 & -38 \end{vmatrix} \end{aligned} \tag{1.5.2}$$

The determinant of the 4×4 matrix is written in terms of the determinants of four 3×3 matrices, one application of recursive formula (1.5.1). To this point, the computation is not so bad. It would take a minute to write down this quantity by hand. However, you might feel no closer to the final result, which is $-393,294$, than before. Now there are four separate determinants to determine. To continue the computation, the determinant of each 3×3 matrix would be written in terms of the determinants of three 2×2 matrices, a second application of formula (1.5.1). Thus the determinant of A would be written in terms of twelve 2×2 determinants. A final application of formula (1.5.1) would yield the determinant of A in terms of twenty-four 1×1 determinants (scalars), at which point the arithmetic could be done and the determinant determined. Hopefully you are convinced that completing this calculation by hand would take a while and be prone to error.

The main point of this discourse is to familiarize you with the recursive definition. Making sure you get the right signs on the coefficients and extract the right submatrices at each step takes some practice. Can you use formula (1.5.1) to find $\det \begin{pmatrix} 2 & 4 \\ -1 & 3 \end{pmatrix}$? Answer on page 32.

The quantities $(-1)^{1+j} \det A_{\setminus 1,j}$ of formula (1.5.1) are called **cofactors**. More generally, the quantity $(-1)^{i+j} \det A_{\setminus i,j}$ is called the i,j-cofactor of A. Cofactors can be computed for any row-column combination. Using the notation $C_{i,j}$ for the i,j-cofactor, recursion (1.5.1) can be rewritten

$$\det A = A_{1,1} C_{1,1} + A_{1,2} C_{1,2} + \cdots + A_{1,n} C_{1,n}. \tag{1.5.3}$$

[4]Formula (1.5.1) can be made to work for 1×1 matrices by defining $\det A_{\setminus 1,1} = 1$ for a 1×1 matrix A.

While more succinct, this presentation hides the details of the calculation. Each $C_{i,j}$ may be an involved calcuation itself.

The expression

$$-12\begin{vmatrix} 45 & -46 & 23 \\ -28 & 4 & -48 \\ 34 & -38 & -18 \end{vmatrix} - 49\begin{vmatrix} 28 & -46 & 23 \\ -15 & 4 & -48 \\ -1 & -38 & -18 \end{vmatrix}$$
$$-45\begin{vmatrix} 28 & 45 & 23 \\ -15 & -28 & -48 \\ -1 & 34 & -18 \end{vmatrix} + 10\begin{vmatrix} 28 & 45 & -46 \\ -15 & -28 & 4 \\ -1 & 34 & -38 \end{vmatrix}$$

from calculation (1.5.2) is an example of a **linear combination**. It is the sum of scalar multiples of matrices. More generally, if S is any set of objects on which addition and scalar mutliplication are defined, $c_1, c_2, \ldots, c_n$ are scalars, and objects $b_1, b_2, \ldots, b_n$ are in S, then the expression

$$c_1 b_1 + c_2 b_2 + \cdots + c_n b_n$$

is called a linear combination of the objects $b_1, b_2, \ldots, b_n$, and $c_1, c_2, \ldots, c_n$ are called the **coefficients** of the linear combination.

Crumpet 9: Linear Combinations

Linear combinations appear in many contexts.

- A polynomial in t is a linear combination of the monomials $1, t, t^2, t^3, \ldots, t^n$.
- A Riemann sum is a linear combination of certain values of a function.
- The solutions of the differential equation $y'' - 4y' + 3y = 0$ are linear combinations of the functions e^x and e^{3x}.
- Numerical approximations of derivatives, such as $-\frac{3}{2h}f(x_0) + \frac{2}{h}f(x_0+h) - \frac{1}{2h}f(x_0+2h)$, are linear combinations of certain values of a function.
- The left-hand side of the equation $3x - 2y = 7$ is a linear combination of the variables x and y.
- The expected value of a random variable with finitely many possible values is a linear combination.

Addition and scalar multiplication are defined for objects such as functions, variables, numbers, integrals, vectors, and matrices. Each of the following is a linear combination.

$$3\sin(x) - 2\sin(2x) + \sin(3x) \qquad 7x + 2y - \frac{4}{5}z$$
$$6\sqrt{2} - 2\sqrt{7} \qquad \int_0^1 f(x)dx + \int_1^2 f(x)dx + \int_2^3 f(x)dx + \int_3^4 f(x)dx$$
$$\frac{1}{\sqrt{5}}\langle -2, 1\rangle - \frac{1}{\sqrt{13}}\langle 3, -2\rangle \qquad 2\begin{bmatrix} 2 & -6 \\ 0 & 3 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 2 & 0 \\ 4 & -3 \end{bmatrix}$$

Can you think of other places where you've seen linear combinations?

Sudoku Row Linear Combinations

If you enjoy solving sudoku puzzles, give this one a shot before reading on. Answer on page 32.

			5	6		1	9	
7						4		5
1	5	6			9		3	
3	7	1	6			9		
9				1				7
		5			2	3	1	6
	2		1			5	8	9
5		7						1
	1	8		4	5			

Can the third row of the 2,1-block of the completed sudoku board be written as a linear combination of its first two rows? Maybe you feel there is a natural way to understand linear combinations of rows of a sudoku puzzle and maybe not. It is not something done in solving the puzzle. However, if we cast each sudoku row as a 1×3 (row) matrix, operate on the row matrices and then cast back to sudoku rows, it would be as if the sudoku rows themselves were being added. In mathematics, we might say the sudoku rows inherit the operations of addition and scalar multiplication from the corresponding operations on matrices.

For example,

1	4	7	+	2	6	3

$\Downarrow$ (casting to matrices)

$$\begin{bmatrix} 1 & 4 & 7 \end{bmatrix} + \begin{bmatrix} 2 & 6 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 10 & 10 \end{bmatrix}$$

(casting to sudoku rows) $\Downarrow$

3	10	10

Scalar multiplication on sudoku rows is inherited in the same manner. With addition and scalar multiplication inherited, linear combinations are inherited. Back to the question...

Can the third row of the 2,1-block of the completed sudoku board (on page 32) be written as a linear combination of the first two rows? Rephrasing, does the following equation have a solution?

a	3	7	1	$+\,b$	9	6	2	=	8	4	5

Casting the equation in terms of matrices and solving:

$$a\begin{bmatrix} 3 & 7 & 1 \end{bmatrix} + b\begin{bmatrix} 9 & 6 & 2 \end{bmatrix} = \begin{bmatrix} 8 & 4 & 5 \end{bmatrix}$$
$$\begin{bmatrix} 3a+9b & 7a+6b & a+2b \end{bmatrix} = \begin{bmatrix} 8 & 4 & 5 \end{bmatrix}$$

For these two row matrices to be equal corresponding entries must be equal. That is, the simultaneous equations

$$\begin{aligned} 3a + 9b &= 8 \\ 7a + 6b &= 4 \\ a + 2b &= 5 \end{aligned}$$

must all be true. The second and third equations can be solved (as a system) by elimination, for example. The second equation minus 3 times the third equation yields $4a = -11$, so $a = \frac{-11}{4}$. Substituting into the third equation yields $\frac{-11}{4} + 2b = 5$ which means $b = \frac{31}{8}$. These values of a and b constitute the only simultaneous solution of the second and third equations. Substituting into the first equation yields $3\left(\frac{-11}{4}\right) + 9\left(\frac{31}{8}\right) = 8$ which can be confirmed FALSE! Therefore there is no solution. There is no way to write the third row of the 2,1-block as a linear combination of the first two rows.

By contrast the third row of the 1,3-block can be written as -1 times the first row plus 2 times the second row. The third row is the linear combination of the first two rows with coefficients -1 and 2. Can you verify this? The 1,3-block is

1	9	8
4	6	5
7	3	2

.

Answer on page 32.

Through the process of inheritance the determinant of any 3×3 sudoku block can also be calculated. For example, the determinant of the 2,1-block is

$$\begin{aligned}
\begin{vmatrix} 3 & 7 & 1 \\ 9 & 6 & 2 \\ 8 & 4 & 5 \end{vmatrix} &= 3\begin{vmatrix} 6 & 2 \\ 4 & 5 \end{vmatrix} - 7\begin{vmatrix} 9 & 2 \\ 8 & 5 \end{vmatrix} + 1\begin{vmatrix} 9 & 6 \\ 8 & 4 \end{vmatrix} \\
&= 3(6 \cdot 5 - 2 \cdot 4) - 7(9 \cdot 5 - 2 \cdot 8) + 1(9 \cdot 4 - 6 \cdot 8) \\
&= 3(22) - 7(29) + 1(-12) \\
&= 66 - 203 - 12 \\
&= -149
\end{aligned}$$

What is the determinant of the 1,3-block? Answer on page 32.

So, for the block with determinant -149 there was no way to write the third row as a linear combination of the first two, and for the block with determinant 0 there was a way to write the third row as a linear combination of the first two. This bears further investigation, requested in the exercises.

Key Concepts

coefficients The scalar quantities of a linear combination.

cofactor A scalar quantity denoted $C_{i,j}$, computed for the matrix A as

$$C_{i,j} = (-1)^{i+j} \det A_{\setminus i,j} \tag{1.5.4}$$

determinant The determinant of an $n \times n$ matrix A, denoted $\det A$ or $|A|$, is defined by

$$\det A = A_{1,1}C_{1,1} + A_{1,2}C_{1,2} + \cdots + A_{1,n}C_{1,n}$$

for $n > 1$ and $\det A = A_{1,1}$ for $n = 1$. The determinant of an $m \times n$ matrix is undefined if $m \neq n$.

linear combination An expression of the form

$$c_1 b_1 + c_2 b_2 + \cdots + c_n b_n = \sum_{i=1}^{n} c_i b_i$$

where $c_1, c_2, \ldots, c_n$ are scalars and $b_1, b_2, \ldots, b_n$ are objects from a set on which addition and scalar multiplication are defined.

square matrix A matrix with the same number or columns as rows. An $n \times n$ matrix.

SageMath

If `M` is a matrix in SageMath, then `M.determinant()` is its determinant. The following code computes the determinant of $A = \begin{bmatrix} -12 & 49 & -45 & -10 \\ 28 & 45 & -46 & 23 \\ -15 & -28 & 4 & -48 \\ -1 & 34 & -38 & -18 \end{bmatrix}$, the matrix behind calculation (1.5.2). SageMathCell [27]

```
M = matrix(4,4,[-12,49,-45,-10,28,45,-46,23,-15,-28,4,-48,-1,34,-38,-18])
print(M.determinant())
```

The output of this code is

```
-393294
```

Exercises

1. Use formula (1.5.4) to write the cofactor as a determinant.

   (a) $C_{1,1}$ of $\begin{bmatrix} -9 & 3 \\ 0 & 6 \end{bmatrix}$

   (b) $C_{1,2}$ of $\begin{bmatrix} 9 & -6 \\ -11 & -9 \end{bmatrix}$

   (c) $C_{2,1}$ of $\begin{bmatrix} -4 & -6 \\ 9 & -11 \end{bmatrix}$ [𝕊]-284

   (d) $C_{2,2}$ of $\begin{bmatrix} 8 & 13 \\ 6 & 9 \end{bmatrix}$ [𝔸]-348

   (e) $C_{1,3}$ of $\begin{bmatrix} -2 & -1 & 6 \\ -5 & -10 & 11 \\ -9 & 8 & 0 \end{bmatrix}$ [𝕊]-284

   (f) $C_{2,1}$ of $\begin{bmatrix} -3 & 0 & 1 \\ 3 & -5 & 7 \\ 4 & 11 & -7 \end{bmatrix}$ [𝔸]-348

   (g) $C_{2,2}$ of $\begin{bmatrix} -12 & -2 & -10 \\ 5 & 0 & 3 \\ 2 & -9 & -4 \end{bmatrix}$

   (h) $C_{3,2}$ of $\begin{bmatrix} 7 & -6 & -7 \\ -2 & -4 & -11 \\ 2 & 0 & 4 \end{bmatrix}$

2. Calculate the determinant if possible.

   (a) $\det A$ where $A = \begin{bmatrix} 30 \end{bmatrix}$

   (b) $|A|$ where $A = \begin{bmatrix} -6 \end{bmatrix}$

   (c) $\det A$ where $A = \begin{bmatrix} -45 \end{bmatrix}$ [𝕊]-284

   (d) $\det A$ where $A = \begin{bmatrix} 44 \end{bmatrix}$ [𝔸]-348

   (e) $\det \begin{pmatrix} 5 & -2 \\ 7 & 2 \end{pmatrix}$

   (f) $\begin{vmatrix} -18 & 19 \\ 6 & -3 \end{vmatrix}$

   (g) $\begin{vmatrix} 18 & 5 \\ 14 & -16 \end{vmatrix}$ [𝕊]-284

   (h) $\det \begin{pmatrix} -11 & 2 \\ -10 & -7 \end{pmatrix}$ [𝔸]-348

   (i) $\begin{vmatrix} -5 & 2 & -4 \\ 9 & 0 & -2 \\ -6 & 8 & 4 \end{vmatrix}$

   (j) $\begin{vmatrix} 0 & 9 & 7 \\ -1 & -6 & -2 \\ 6 & -9 & -5 \end{vmatrix}$

   (k) $\det \begin{pmatrix} -3 & -1 & -9 \\ 1 & -4 & -8 \\ 2 & 9 & 6 \end{pmatrix}$ [𝕊]-284

   (l) $\begin{vmatrix} 3 & -6 & 0 \\ -8 & 2 & -7 \\ 5 & 1 & -1 \end{vmatrix}$ [𝔸]-348

   (m) $\det \begin{pmatrix} 2 & 8 & -2 & 0 \\ 3 & 8 & 1 & 2 \\ 0 & 0 & 1 & 6 \\ 2 & 0 & -1 & 6 \end{pmatrix}$

   (n) $\begin{vmatrix} 4 & 5 & -2 & 0 \\ 2 & 0 & 4 & 7 \\ 8 & 4 & 5 & -2 \\ 2 & -2 & 0 & 0 \end{vmatrix}$

   (o) $\det \begin{pmatrix} 5 & 0 & 2 & 8 \\ 4 & 8 & 6 & -2 \\ 0 & -1 & 6 & 0 \\ 0 & 3 & -1 & 3 \end{pmatrix}$ [𝕊]-284

3. Formula (1.5.1) reduces the calculation of the determinant of a 4×4 matrix into a linear combination of the determinants of twenty-four 1×1 matrices, as in calculation (1.5.2).

   (a) Formula (1.5.1) reduces the calculation of the determinant of a 5×5 matrix into a linear combination of the determinants of how many 1×1 matrices?

   (b) Formula (1.5.1) reduces the calculation of the determinant of an $n \times n$ matrix into a linear combination of the determinants of how many 1×1 matrices?

4. True or false? [𝔸]-348

   (a) The determinant of a matrix is a scalar.

   (b) The determinant of a matrix is always positive since it is the absolute value of a number.

   (c) The determinant of a 5×6 matrix can be written as a linear combination of the determinants of thirty 1×1 matrices.

   (d) If A and B are 1×1 matrices, then $\det A + \det B = \det(A + B)$.

   (e) If A and B are 2×2 matrices, then $\det A + \det B = \det(A + B)$.

5. Compare and contrast (i) scalar, (ii) 1×1 matrix, and (iii) the determinant of a 1×1 matrix.

6. Calculate the determinant. **HINT**: Despite the large sizes of some of the matrices, this does not require a lot of work.

   (a) $\begin{vmatrix} 7 & 0 \\ 6 & -8 \end{vmatrix}$

   (b) $\begin{vmatrix} -2 & 0 & 0 \\ -7 & 4 & 0 \\ 2 & -9 & 7 \end{vmatrix}$

   (c) $\begin{vmatrix} -4 & 0 & 0 & 0 \\ -8 & -9 & 0 & 0 \\ 2 & 7 & 6 & 0 \\ -5 & -7 & 3 & 9 \end{vmatrix}$

   (d) $\begin{vmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ -7 & -3 & 0 & 0 & 0 & 0 \\ -6 & -4 & 3 & 0 & 0 & 0 \\ -1 & 5 & 1 & -2 & 0 & 0 \\ 4 & -5 & 8 & 3 & 4 & 0 \\ 8 & 9 & 7 & -9 & 0 & 9 \end{vmatrix}$

7. In your own words, draw a conjecture based on the calculations of question 6.
8. SageMathCell 28 Calculate
   (a) $\det A$
   (b) $\det B$
   (c) $\det(AB)$
   (d) $\det(3A)$
   (e) $\det(3B)$

   What do you notice?

The remaining exercises refer to the completed sudoku board of this section (page 32).

9. SageMathCell 29 The determinants of the 1,3-block and the 2,1-block are 0 and -149 respectively. Find the determinants of the remaining 7 blocks.
10. SageMathCell 30 For the 1,3-block, the third row can be written as a linear combination of the first two (-1 times the first row plus 2 times the second row). For the 2,1-block, the third row cannot be written as a linear combination of the first two. For the remaning 7 blocks, explore whether there is any row that can be written as a linear combination of the other two. [A]-348
11. Make a conjecture about the connection between determinant and the possibility of writing one of the rows of a block as a linear combination of the others.
12. SageMathCell 31 Can any of the 9 rows of the sudoku board be written as a linear combination of the other 8? Apply your conjecture from question 11 to answer the question.

Answers

determinant:

$$\begin{aligned}\det\begin{pmatrix} 2 & 4 \\ -1 & 3 \end{pmatrix} &= (-1)^{1+1}(2)\det(3) + (-1)^{1+2}(4)\det(-1) \\ &= 2(3) - 4(-1) \\ &= 10\end{aligned}$$

sudoku:

2	3	4	5	6	7	1	9	8
7	8	9	2	3	1	4	6	5
1	5	6	4	8	9	7	3	2
3	7	1	6	5	8	9	2	4
9	6	2	3	1	4	8	5	7
8	4	5	7	9	2	3	1	6
4	2	3	1	7	6	5	8	9
5	9	7	8	2	3	6	4	1
6	1	8	9	4	5	2	7	3

linear combination:

$$\begin{aligned}(-1)\begin{bmatrix} 1 & 9 & 8 \end{bmatrix} + 2\begin{bmatrix} 4 & 6 & 5 \end{bmatrix} &= \begin{bmatrix} -1+8 & -9+12 & -8+10 \end{bmatrix} \\ &= \begin{bmatrix} 7 & 3 & 2 \end{bmatrix}\end{aligned}$$

sudoku determinant:

$$\begin{aligned}\begin{vmatrix} 1 & 9 & 8 \\ 4 & 6 & 5 \\ 7 & 3 & 2 \end{vmatrix} &= 1\begin{vmatrix} 6 & 5 \\ 3 & 2 \end{vmatrix} - 9\begin{vmatrix} 4 & 5 \\ 7 & 2 \end{vmatrix} + 8\begin{vmatrix} 4 & 6 \\ 7 & 3 \end{vmatrix} \\ &= 1(6\cdot 2 - 5\cdot 3) - 9(4\cdot 2 - 5\cdot 7) + 8(4\cdot 3 - 6\cdot 7) \\ &= 1(-3) - 9(-27) + 8(-30) \\ &= -3 + 243 - 240 \\ &= 0\end{aligned}$$

1.6 Matrix "Division"

Given a message such as "Hello World!", a system for converting letters and symbols to numbers can be used to turn the message into a list of numbers. These numbers can be further disguised by multiplying by an encoding matrix, giving a new list of numbers, an secret message! The following list of numbers was created this way.

-199 -78 14 -273 -145 -13 -572 -294 -49 -245 -127 -7 -150 -84 -1
-389 -174 -10 412 272 103 -142 -59 16 -231 -132 -33

Can you decode it? Learning how to decode a message like this is the topic of this section.

You may have heard the claim "there's no such thing as subtraction—it's just adding the opposite" or something like it. There is a vital concept of linear algebra buried in this addage. The link between addition, opposites, and zero that makes subtraction optional is a well known property of real numbers. The sum of opposites is zero.

Why zero, and not some other number? Zero is that special number that can be added to any number without changing its value. There is no other! In symbols, $a + 0 = 0 + a = a$. This property is so special it has a name. Zero is the **additive identity** for real numbers—the word *identity* to signal this special property and the word *additive* to document the operation. The additive inverse, or opposite, of a real number ***is defined*** by the fact that adding the two yields the additive identity. Two numbers are additive inverses (opposites) if and only if their sum is the additive identity (zero).

Likewise, one is that special number that can be multiplied by any number without changing its value. Consequently, one is the **multiplicative identity** for real numbers. In symbols, $a \cdot 1 = 1 \cdot a = a$. The multiplicative inverse (reciprocal) of a real number ***is defined*** by the fact that multiplying the two yields the multiplicative identity. Two numbers are multiplicative inverses (reciprocals) if and only if their product is the multiplicative identity (one).

The link between multiplication, reciprocals, and one is analogous to the link between addition, opposites, and zero. For any real numbers a and b,

$$a \text{ and } b \text{ are reciprocals if and only if } a \cdot b = b \cdot a = 1$$
$$\text{and}$$
$$a \text{ and } b \text{ are opposites if and only if } a + b = b + a = 0.$$

The same relationship holds among addition, opposites, and zero as holds among multiplication, reciprocals, and one. Addition and multiplication are operations, opposites and reciprocals are inverses, and zero and one are identities.

There is an important analogy for matrices. To see it, compute the following products.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ -8 \end{bmatrix} \qquad \begin{bmatrix} 3 & 4 & -1 \\ 17 & -21 & 55 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{5} & -\pi & 18 & \frac{2}{3} \\ \frac{17}{8} & 34 & \sqrt[5]{19} & \ln 9 \\ \frac{\pi}{4} & 0.34 & e^7 & \sin(1) \\ 2^{2^2} & \tan^{-1}(1) & 12 & 10^{1000} \end{bmatrix}$$

Answers on page 39. Hopefully these exercises have led you to the conclusion that multiplying by matrices such as

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

leaves the multiplicand (the matrix being multiplied by it) unchanged. Multiplying a matrix by matrices such as these does not change the value of the matrix, the exact same property that made 1 the multiplicative identity and 0 the additive identity for real numbers. By extension, that makes these matrices identity matrices. They can each be multiplied with any other matrix (as long as the product is defined) without changing the other's value.

The $n \times n$ **identity matrix** is denoted $I_{n\times n}$ or just I when the size of the matrix is known or unimportant. Identity matrices have ones on the main diagonal, the diagonal running from the 1,1-entry through the n,n-entry, and zeros elsewhere. In symbols, $I_{1,1} = I_{2,2} = \cdots = I_{n,n} = 1$ and $I_{j,k} = 0$ whenever $j \neq k$. For any matrix M, $M \cdot I = I \cdot M = M$.

With an identity, or really set of identities, for multiplication, we are only one element shy of the operation, inverse, identity triumvirate for matrices—the inverse. Compute the following products.

$$\begin{bmatrix} 7 & 4 \\ 5 & 3 \end{bmatrix}\begin{bmatrix} 3 & -4 \\ -5 & 7 \end{bmatrix} \qquad \begin{bmatrix} -4 & 5 & 3 \\ 8 & -1 & -2 \\ -3 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & -2 & -7 \\ -2 & 5 & 16 \\ 5 & -11 & -36 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 0 & -3 & -2 \\ 1 & 1 & -6 & -3 \\ -2 & -3 & 10 & 5 \\ 0 & 0 & 3 & 1 \end{bmatrix}\begin{bmatrix} 4 & -9 & -3 & -4 \\ -1 & 1 & 0 & 1 \\ -1 & 3 & 1 & 2 \\ 3 & -9 & -3 & -5 \end{bmatrix}$$

Answers on page 40. These exercises demonstrate that there are pairs of matrices A and B such that $AB = I$. But what about BA? In our formulations for real number inverses we had $a+b = b+a = 0$ and $a \cdot b = b \cdot a = 1$. Unfortunately we observed in section 1.3 that matrix multiplication is not commutative. We cannot immediately conclude that $BA = I$ just because $AB = I$. Will we get lucky, though?

Compute the following products (the same as above only in the opposite order).

$$\begin{bmatrix} 3 & -4 \\ -5 & 7 \end{bmatrix}\begin{bmatrix} 7 & 4 \\ 5 & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & -2 & -7 \\ -2 & 5 & 16 \\ 5 & -11 & -36 \end{bmatrix}\begin{bmatrix} -4 & 5 & 3 \\ 8 & -1 & -2 \\ -3 & 1 & 1 \end{bmatrix}$$
$$\begin{bmatrix} 4 & -9 & -3 & -4 \\ -1 & 1 & 0 & 1 \\ -1 & 3 & 1 & 2 \\ 3 & -9 & -3 & -5 \end{bmatrix}\begin{bmatrix} 1 & 0 & -3 & -2 \\ 1 & 1 & -6 & -3 \\ -2 & -3 & 10 & 5 \\ 0 & 0 & 3 & 1 \end{bmatrix}$$

As you were hopefully able to verify, all of these products are identity matrices too! It seems that multiplicative inverse pairs commute. That is, if $AB = I$, then $BA = I$. We finally have evidence that a matrix analogy for linking multiplication, inverses, and identity matrices exists.

For any matrices A and B,

$$A \text{ and } B \text{ are inverses if and only if } A \cdot B = B \cdot A = I. \tag{1.6.1}$$

Crumpet 10: Inverses of non-square matrices?

Suppose A is an $m \times n$ matrix and there are matrices L and R such that $LA = I$ and $AR = I$. By theorems 5 and 6 A must have a pivot position in every row and every column (see section 2.2 for a definition of pivot position). The only way that can happen is if A is square. Hence a matrix with left and right inverses must be square.

This is enough that we could define $\frac{I}{A}$ as the multiplicative inverse (or reciprocal) of A and have the understanding that $A \div B$ means $A \cdot \frac{I}{B}$ much like we have for real numbers, but by convention we do not! For one thing, we would need a second division symbol to mean $\frac{I}{B} \cdot A$ since matrix multiplication is not commutative. In general, $\frac{I}{B} \cdot A$ and $A \cdot \frac{I}{B}$ could be unequal. Instead, we stick with the addage that "there's no such thing as matrix division—it's just multiplying by the inverse". The notation we use for the inverse of A is A^{-1}, borrowing from the algebra of real numbers but not using division bars or division symbols.

A Formula for the Inverse (proven in section 3.7)

For any matrix A, if A is invertible then

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} C_{1,1} & C_{2,1} & \cdots & C_{m,1} \\ C_{1,2} & C_{2,2} & \cdots & C_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1,n} & C_{2,n} & \cdots & C_{m,n} \end{bmatrix} \tag{1.6.2}$$

where the $C_{i,j}$ are the cofactors of A. This implies that when A^{-1} exists

1. A must be square since $\det A$ and the $C_{i,j}$ are undefined if A is not square, and
2. $\det A$ must be nonzero since division by 0 is undefined.

When A^{-1} is defined, we say that A is **invertible**. The matrix

$$\begin{bmatrix} C_{1,1} & C_{2,1} & \cdots & C_{n,1} \\ C_{1,2} & C_{2,2} & \cdots & C_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1,n} & C_{2,n} & \cdots & C_{n,n} \end{bmatrix}$$

is called the **adjugate** of A, adj$A$. With this definition, the formula for the inverse can be summarized as

$$A^{-1} = \frac{1}{\det A}\operatorname{adj}A.$$

One Property of the Inverse

Multiplication by a matrix's inverse "undoes" multiplication by the matrix just as dividing by a number undoes multiplication by that same number. In symbols, if A and B are matrices and B is invertible (has an inverse)

$$(AB)B^{-1} = A \tag{1.6.3}$$

much like $(a \cdot b) \div b = a$ for real numbers. If we used division in linear algebra, the equation $(AB)B^{-1} = A$ might be written $(A \cdot B) \div B = A$, making the comparison clearer. The only potential harm in thinking with division is that $(BA)B^{-1}$ is generally not A, so $(B \cdot A) \div B \neq A$ for matrices even though $(b \cdot a) \div b = a$ for real numbers. Since multiplication of matrices is not commutative, right-multiplication by B^{-1} does not undo left-multiplication by B. Using the notation B^{-1} and paying close attention to right-multiplication versus left-multiplication will help keep this straight.

To illustrate, let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -4 \\ -5 & 7 \end{bmatrix}$, making

$$AB = \begin{bmatrix} -7 & 10 \\ -11 & 16 \end{bmatrix}.$$

Can you verify this? As seen earlier, $B^{-1} = \begin{bmatrix} 7 & 4 \\ 5 & 3 \end{bmatrix}$. Compute $(AB)B^{-1}$ to see that it equals A, and compute $B^{-1}(AB)$ to see that it does not equal A. Answer on page 40.

Inverses and Cryptography

The ability to undo multiplication by an invertible matrix makes it possible to use matrices and their inverses for encrypting and decrypting messages. Decoding messages like the one opening this section amounts to regrouping the code into column vectors and multiplying by the decoding matrix—the inverse of the coding matrix. As long as the parties on either end of the message transmission have one matrix of some pair of inverse matrices, they can each encode with their matrix, send their message securely, and decode received messages. Without knowledge of the coding or decoding matrix, an intercepted message would be very difficult to decode!

Table 1.2: ASCII (American Standard Code for Information Interchange) characters

Dec	Hex	Char		Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char
0	0	NUL	(null character)	32	20	(space)	64	40	@	96	60	`
1	1	SOH	(start of heading)	33	21	!	65	41	A	97	61	a
2	2	STX	(start of text)	34	22	"	66	42	B	98	62	b
3	3	ETX	(end of text)	35	23	#	67	43	C	99	63	c
4	4	EOT	(end of transmission)	36	24	$	68	44	D	100	64	d
5	5	ENQ	(enquiry)	37	25	%	69	45	E	101	65	e
6	6	ACK	(acknowledge)	38	26	&	70	46	F	102	66	f
7	7	BEL	(bell)	39	27	'	71	47	G	103	67	g
8	8	BS	(backspace)	40	28	(	72	48	H	104	68	h
9	9	HT	(horizontal tab)	41	29	)	73	49	I	105	69	i
10	A	LF	(line feed)	42	2A	*	74	4A	J	106	6A	j
11	B	VT	(vertical tab)	43	2B	+	75	4B	K	107	6B	k
12	C	FF	(form feed)	44	2C	,	76	4C	L	108	6C	l
13	D	CR	(carriage return)	45	2D	-	77	4D	M	109	6D	m
14	E	SO	(shift out)	46	2E	.	78	4E	N	110	6E	n
15	F	SI	(shift in)	47	2F	/	79	4F	O	111	6F	o
16	10	DLE	(data link escape)	48	30	0	80	50	P	112	70	p
17	11	DC1	(device control 1)	49	31	1	81	51	Q	113	71	q
18	12	DC2	(device control 2)	50	32	2	82	52	R	114	72	r
19	13	DC3	(device control 3)	51	33	3	83	53	S	115	73	s
20	14	DC4	(device control 4)	52	34	4	84	54	T	116	74	t
21	15	NAK	(negative acknowledge)	53	35	5	85	55	U	117	75	u
22	16	SYN	(synchronous idle)	54	36	6	86	56	V	118	76	v
23	17	ETB	(end of transmission block)	55	37	7	87	57	W	119	77	w
24	18	CAN	(cancel)	56	38	8	88	58	X	120	78	x
25	19	EM	(end of medium)	57	39	9	89	59	Y	121	79	y
26	1A	SUB	(substitute)	58	3A	:	90	5A	Z	122	7A	z
27	1B	ESC	(escape)	59	3B	;	91	5B	[	123	7B	{
28	1C	FS	(file separator)	60	3C	<	92	5C	\	124	7C	|
29	1D	GS	(group separator)	61	3D	=	93	5D	]	125	7D	}
30	1E	RS	(record separator)	62	3E	>	94	5E	^	126	7E	~
31	1F	US	(unit separator)	63	3F	?	95	5F	_	127	7F	(delete)

All we need now is a method for converting letters and symbols to numbers and numbers back to letters and symbols. While a basic conversion from letters to numbers would have each letter of the alphabet assigned a number from 1-26 or 0-25, this would leave punctuation, symbols like spaces and hashtags, capitalization, and numbers out. Since the early 1960's the American National Standards Institute has maintained a coding system for the electronic transmission of documents in English called ASCII (pronounced ass-kee) or US-ASCII. Part of that system, largely developed by Bob Bemer [18], is a numeric representation of all the symbols you are likely to find on an English language keyboard. See Table 36. For example, the capital letter "A" has numeric code 65, the lower case letter "a" has numeric code 97, and the space has numeric code 32.

Using the coding matrix $\begin{bmatrix} -7 & 3 & 2 \\ -4 & 1 & 2 \\ -1 & 0 & 1 \end{bmatrix}$, the message "Hello World!" would be encrypted as follows.

1. "Hello World!" is converted to the numeric sequence 72 101 108 108 111 32 87 111 114 108 100 33 using ASCII.

2. Since we are using a 3×3 coding matrix, the numeric sequence is grouped three at a time into the 3-row matrix

$$\begin{bmatrix} 72 & 108 & 87 & 108 \\ 101 & 111 & 111 & 100 \\ 108 & 32 & 114 & 33 \end{bmatrix}$$

   If the message did not have a multiple of three characters, 0's (null characters) could be added to the end.

3. The message matrix is multiplied by the coding matrix.

$$\begin{bmatrix} -7 & 3 & 2 \\ -4 & 1 & 2 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 72 & 108 & 87 & 108 \\ 101 & 111 & 111 & 100 \\ 108 & 32 & 114 & 33 \end{bmatrix} = \begin{bmatrix} 15 & -359 & -48 & -390 \\ 29 & -257 & -9 & -266 \\ 36 & -76 & 27 & -75 \end{bmatrix}$$

   This is a good place to use a calculator or SageMath! SageMath Cell 32

4. The coded message is extracted from the product:

   15 29 36 -359 -257 -76 -48 -9 27 -390 -266 -75

The message at the beginning of this section was encoded with the same matrix. Can you decode it (using a calculator or SageMath to assist)? Answer on page 40.

Crumpet 11: Lester S. Hill

The first documented multiple-letter cipher is attributed to Lester S. Hill. His *Mathematical Monthly* article of 1929 [12] outlines a procedure very similar to the one presented here except modular arithmetic is used to make sure all numbers in the encoded message are valid character codes. Thus the encoded message is transmitted as a sequence of letters and symbols, not numbers. His work far predates electronic computing devices so, to be practical, he needed a way to limit the difficulty of doing the computations, a second impetus for using modular arithmetic.

Key Concepts

A^{-1} The inverse of A. Can be computed via (1.6.2).

adjugate For a square matrix, the transpose of its matrix of cofactors.

identity matrix A matrix with ones on the main diagonal and zeros elsewhere.

matrix inverse Matrices A and B are inverses of one another if and only if $AB = BA = I$.

invertible (matrix) A matrix whose inverse is defined.

main diagonal The i,i-entries of a matrix.

SageMath

If `M` is a matrix in SageMath, then `M.inverse()` is its inverse. The following code computes the inverses of $A = \begin{bmatrix} 4 & -9 & -3 & -4 \\ -1 & 1 & 0 & 1 \\ -1 & 3 & 1 & 2 \\ 3 & -9 & -3 & -5 \end{bmatrix}$ and $B = \begin{bmatrix} \sqrt{5} & -\pi & 18 \\ \frac{17}{8} & 34 & \sqrt[5]{19} \\ \frac{\pi}{4} & 0.34 & e^7 \end{bmatrix}$.

```
A = matrix(4,4,[4,-9,-3,-4,-1,1,0,1,-1,3,1,2,3,-9,-3,-5])
print(A.inverse()); print()
B=matrix(3,3,[sqrt(5),-pi,18,17/8,34,19^(1/5),pi/4,0.34,e^7])
print(B.inverse())
```

SageMathCell 33 The output for A^{-1} is

```
[ 1  0 -3 -2]
[ 1  1 -6 -3]
[-2 -3 10  5]
[ 0  0  3  1]
```

but the output for B^{-1} is far too long to fit on the page. Just the 1,1-entry is $1/5(\sqrt{5}\pi - \sqrt{5}(\sqrt{5}\pi^2 + 6.8)/(\sqrt{5}\pi + 80))(\sqrt{5}\pi(153\sqrt{5} - 20 \cdot 19^{1/5}))/(\sqrt{5}\pi + 80) - 153\sqrt{5})/(153\sqrt{5}\pi - (\sqrt{5}\pi^2 + 6.8)(153\sqrt{5} - 20 \cdot 19^{1/5})/(\sqrt{5}\pi + 80) - 170e^7) + 1/5\sqrt{5} - \pi/(\sqrt{5}\pi + 80)$.

Exercises

1. Compute the inverse if possible.

   (a) $\begin{bmatrix} -3 \end{bmatrix}$

   (b) $\begin{bmatrix} 0 \end{bmatrix}$

   (c) $\begin{bmatrix} \frac{2}{\sqrt{3}} \end{bmatrix}$ [𝕊]-285

   (d) $\begin{bmatrix} 4\pi \end{bmatrix}$ [𝔸]-348

   (e) $\begin{bmatrix} 3 & -2 \\ 11 & -7 \end{bmatrix}$

   (f) $\begin{bmatrix} \sqrt{12} & 3 \\ 2 & \sqrt{3} \end{bmatrix}$

   (g) $\begin{bmatrix} 5 & \sqrt{18} \\ \sqrt{8} & 3 \end{bmatrix}$ [𝔸]-348

   (h) $\begin{bmatrix} 5 & -3 \\ -5 & 4 \end{bmatrix}$ [𝕊]-285

   (i) $\begin{bmatrix} 2 & -3 & \sqrt{7} \\ 12 & \sqrt{28} & 5 \end{bmatrix}$ [𝕊]-285

   (j) $\begin{bmatrix} 3 & 4 \\ -7 & 8 \\ -1 & 9 \end{bmatrix}$

   (k) $\begin{bmatrix} 2 & 0 & -2 \\ 0 & 2 & 0 \\ 7 & 8 & 1 \end{bmatrix}$

   (l) $\begin{bmatrix} 0 & -3 & 2 \\ 5 & 1 & 2 \\ 5 & -8 & 8 \end{bmatrix}$

   (m) $\begin{bmatrix} 6 & 3 & 0 \\ -1 & -1 & 6 \\ 0 & 0 & 7 \end{bmatrix}$ [𝕊]-285

   (n) $\begin{bmatrix} 3 & -2 & -6 \\ -1 & 1 & 3 \\ -4 & 3 & 10 \end{bmatrix}$ [𝔸]-348

   (o) $\begin{bmatrix} 9 & -7 & -2 & -11 \\ 0 & 12 & 2 & -12 \\ -1 & -9 & 6 & 10 \end{bmatrix}$

   (p) $\begin{bmatrix} -4 & -9 & 0 \\ 1 & -8 & -1 \\ -10 & 11 & 9 \\ -3 & 8 & 10 \end{bmatrix}$ [𝔸]-348

   (q) $\begin{bmatrix} 2 & 1 & -6 & 2 \\ 1 & 1 & -3 & 0 \\ -4 & -1 & 13 & -7 \\ 3 & 0 & -11 & 9 \end{bmatrix}$ [𝔸]-348

   (r) $\begin{bmatrix} 2 & 0 & 0 & -1 \\ 0 & 1 & -2 & 3 \\ 1 & 0 & 2 & 0 \\ 3 & 1 & 7 & 0 \end{bmatrix}$

(s) $\begin{bmatrix} 4 & 7 & 0 & 8 \\ 0 & 8 & 0 & -1 \\ 0 & 2 & 1 & -1 \\ 0 & 0 & 2 & -2 \end{bmatrix}$

2. Compare and contrast the inverse of a 1×1 matrix with the multiplicative inverse of a real number.

3. True or false? If all the entries in a square matrix M are integers and $\det M = 1$, then all the entries in M^{-1} are integers.

4. Explain how the determinant can help determine whether a matrix has an inverse.

5. Suppose B is an invertible 3×3 matrix and

$$\begin{bmatrix} 1.4 & -70 \\ -29 & 95 \\ -12 & -43 \end{bmatrix} \cdot B = \begin{bmatrix} 80 & 4.9 \\ -62 & -52 \\ -32 & 52 \end{bmatrix}.$$

Find $\begin{bmatrix} 80 & 4.9 \\ -62 & -52 \\ -32 & 52 \end{bmatrix} \cdot B^{-1}$. [𝕊]-286

6. Which matrix is the inverse of

$$\begin{bmatrix} 0 & 0 & 4 & 0 & -1 \\ -2 & 4 & 4 & -1 & 1 \\ -4 & 7 & 10 & -2 & 2 \\ -1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}?$$

(a)
$$\begin{bmatrix} -1 & 4 & -2 & -1 & 8 \\ 0 & 2 & -1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 2 & 0 \\ -1 & 0 & 0 & 0 & 4 \end{bmatrix}$$

(b)
$$\begin{bmatrix} -1 & 4 & -2 & -1 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & -1 & 0 & 2 \\ -1 & 0 & 0 & 4 \end{bmatrix}$$

(c)
$$\begin{bmatrix} 4 & -1 & -2 & -1 & 8 \\ 2 & 0 & -1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 2 & 0 \\ -1 & 0 & 4 & 0 & 0 \end{bmatrix}$$

(d)
$$\begin{bmatrix} 0 & -2 & -4 & -1 & 0 \\ 0 & 4 & 7 & 2 & 0 \\ 4 & 4 & 10 & 0 & 1 \\ 0 & -1 & -2 & 0 & 0 \\ -1 & 1 & 2 & 1 & 0 \end{bmatrix}$$

7. Find x and y so that A and B are inverses.

$$A = \begin{bmatrix} 1 & 1 & -5 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 16 & 11 & 9 \\ 20 & 14 & y \\ x & 5 & 4 \end{bmatrix}$$

HINT: You do not need to calculate A^{-1}. Use the fact that whenever A and B are inverses, $AB = BA = I$.

For the remaining exercises, let $A = \begin{bmatrix} 7 & -5 & -2 \\ -3 & 3 & 1 \\ -3 & 2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 8 & 7 \\ -3 & -5 & -5 \end{bmatrix}$.

8. 34 Compute

(a) $A^{-1}(AB)$

(b) $(AB)A^{-1}$

(c) $B^{-1}(BA)$

(d) $(BA)B^{-1}$

Did you get what you expected?

9. 35 Compute

(a) $(AB)^{-1}$

(b) $A^{-1}B^{-1}$

(c) $B^{-1}A^{-1}$

What do you notice? [𝕊]-286

10. Decode the message -589 861 339 -602 958 317 -244 224 180 -546 768 325 33 -99 0. It was encoded using

$$\begin{bmatrix} 1 & -4 & -2 \\ -3 & 7 & 3 \\ 0 & 2 & 1 \end{bmatrix}.$$

[𝕊]-287

Answers

matrix products part 1:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ -8 \end{bmatrix} = \begin{bmatrix} 3 \\ -8 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 4 & -1 \\ 17 & -21 & 55 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 4 & -1 \\ 17 & -21 & 55 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{5} & -\pi & 18 & \frac{2}{3} \\ \frac{17}{8} & 34 & \sqrt[5]{19} & \ln 9 \\ \frac{\pi}{4} & 0.34 & e^7 & \sin(1) \\ 2^{2^2} & \tan^{-1}(1) & 12 & 10^{1000} \end{bmatrix} = \begin{bmatrix} \sqrt{5} & -\pi & 18 & \frac{2}{3} \\ \frac{17}{8} & 34 & \sqrt[5]{19} & \ln 9 \\ \frac{\pi}{4} & 0.34 & e^7 & \sin(1) \\ 2^{2^2} & \tan^{-1}(1) & 12 & 10^{1000} \end{bmatrix}$$

matrix products part 2:

$$\begin{bmatrix} 7 & 4 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} 3 & -4 \\ -5 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -4 & 5 & 3 \\ 8 & -1 & -2 \\ -3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & -7 \\ -2 & 5 & 16 \\ 5 & -11 & -36 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & -3 & -2 \\ 1 & 1 & -6 & -3 \\ -2 & -3 & 10 & 5 \\ 0 & 0 & 3 & 1 \end{bmatrix} \begin{bmatrix} 4 & -9 & -3 & -4 \\ -1 & 1 & 0 & 1 \\ -1 & 3 & 1 & 2 \\ 3 & -9 & -3 & -5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

inverse undoes multiplication:

(i)

$$(AB)B^{-1} = \begin{bmatrix} -7 & 10 \\ -11 & 16 \end{bmatrix} \begin{bmatrix} 7 & 4 \\ 5 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = A$$

$$((AB)B^{-1})_{1,1} = (-7)(7) + (10)(5) = -49 + 50 = 1$$

$$((AB)B^{-1})_{1,2} = (-7)(4) + (10)(3) = -28 + 30 = 2$$

$$((AB)B^{-1})_{2,1} = (-11)(7) + (16)(5) = -77 + 80 = 3$$

$$((AB)B^{-1})_{2,2} = (-11)(4) + (16)(3) = -44 + 48 = 4$$

(ii)

$$B^{-1}(AB) = \begin{bmatrix} 7 & 4 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} -7 & 10 \\ -11 & 16 \end{bmatrix} = \begin{bmatrix} -93 & 134 \\ -68 & 98 \end{bmatrix} \neq A$$

$$(B^{-1}(AB))_{1,1} = (7)(-7) + (4)(-11) = -49 - 44 = -93$$

$$(B^{-1}(AB))_{1,2} = (7)(10) + (4)(16) = 70 + 64 = 134$$

$$(B^{-1}(AB))_{2,1} = (5)(-7) + (3)(-11) = -35 - 33 = -68$$

$$(B^{-1}(AB))_{2,2} = (5)(10) + (3)(16) = 50 + 48 = 98$$

decoding: The coding matrix $C = \begin{bmatrix} -7 & 3 & 2 \\ -4 & 1 & 2 \\ -1 & 0 & 1 \end{bmatrix}$ has determinant one:

$$\begin{aligned} \det C &= -7\begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} - 3\begin{vmatrix} -4 & 2 \\ -1 & 1 \end{vmatrix} + 2\begin{vmatrix} -4 & 1 \\ -1 & 0 \end{vmatrix} \\ &= -7(1) - 3(-4 + 2) + 2(0 + 1) \\ &= -7 + 6 + 2 \\ &= 1 \end{aligned}$$

The following are cofactors, not entries:

$$C_{1,1} = \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} = 1 \qquad C_{1,2} = -1\begin{vmatrix} -4 & 2 \\ -1 & 1 \end{vmatrix} = 2 \qquad C_{1,3} = \begin{vmatrix} -4 & 1 \\ -1 & 0 \end{vmatrix} = 1$$

$$C_{2,1} = -1\begin{vmatrix} 3 & 2 \\ 0 & 1 \end{vmatrix} = -3 \qquad C_{2,2} = \begin{vmatrix} -7 & 2 \\ -1 & 1 \end{vmatrix} = -5 \qquad C_{2,3} = -1\begin{vmatrix} -7 & 3 \\ -1 & 0 \end{vmatrix} = -3$$

$$C_{3,1} = \begin{vmatrix} 3 & 2 \\ 1 & 2 \end{vmatrix} = 4 \qquad C_{3,2} = -1\begin{vmatrix} -7 & 2 \\ -4 & 2 \end{vmatrix} = 6 \qquad C_{3,3} = \begin{vmatrix} -7 & 3 \\ -4 & 1 \end{vmatrix} = 5$$

so

$$C^{-1} = \frac{1}{1}\mathrm{adj}C = \begin{bmatrix} 1 & -3 & 4 \\ 2 & -5 & 6 \\ 1 & -3 & 5 \end{bmatrix}.$$

Decoding is therefore done by multiplying

$$\begin{bmatrix} 1 & -3 & 4 \\ 2 & -5 & 6 \\ 1 & -3 & 5 \end{bmatrix}\begin{bmatrix} -199 & -273 & -572 & -245 & -150 & -389 & 412 & -142 & -231 \\ -78 & -145 & -294 & -127 & -84 & -174 & 272 & -59 & -132 \\ 14 & -13 & -49 & -7 & -1 & -10 & 103 & 16 & -33 \end{bmatrix}$$
$$= \begin{bmatrix} 91 & 110 & 114 & 108 & 98 & 93 & 8 & 99 & 33 \\ 76 & 101 & 32 & 103 & 114 & 32 & 82 & 107 & 0 \\ 105 & 97 & 65 & 101 & 97 & 83 & 111 & 115 & 0 \end{bmatrix}$$

SageMathCell 36 and the numeric message is 91 76 105 110 101 97 114 32 65 108 103 101 98 114 97 93 32 83 8 82 111 99 107 115 33 0 0. The last step is to look these numbers up in the ASCII table.

1.7 Eigenpairs

Let $A = \begin{bmatrix} -2 & 4 \\ 1 & 1 \end{bmatrix}$, and compute each product before reading on.

$$A\begin{bmatrix} 4 \\ -1 \end{bmatrix} \qquad A\begin{bmatrix} -\frac{8}{3} \\ \frac{2}{3} \end{bmatrix}$$
$$A\begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad A\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$$

You should find that $A\begin{bmatrix} 4 \\ -1 \end{bmatrix} = \begin{bmatrix} -12 \\ 3 \end{bmatrix}$, $A\begin{bmatrix} -\frac{8}{3} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$, $A\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, and $A\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$. Put another way,

$$A\begin{bmatrix} 4 \\ -1 \end{bmatrix} = -3\begin{bmatrix} 4 \\ -1 \end{bmatrix} \qquad A\begin{bmatrix} -\frac{8}{3} \\ \frac{2}{3} \end{bmatrix} = -3\begin{bmatrix} -\frac{8}{3} \\ \frac{2}{3} \end{bmatrix}$$
$$A\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad A\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} = 2\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$$

Put yet another way,

$$A\mathbf{v}_1 = -3\mathbf{v}_1 \qquad A\left(-\frac{2}{3}\mathbf{v}_1\right) = -3\left(-\frac{2}{3}\mathbf{v}_1\right)$$
$$A\mathbf{v}_2 = 2\mathbf{v}_2 \qquad A\left(-\frac{1}{2}\mathbf{v}_2\right) = 2\left(-\frac{1}{2}\mathbf{v}_2\right)$$

where $\mathbf{v}_1 = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, making the relationship between the matrix A and the vectors more apparent. The product of A with each of these vectors gives a scalar multiple of the vector. That's unusual, and a quick experiment will illustrate. Try this:

1. Write down a 2×2 matrix M with four random nonzero entries (from -10 to 10, say).
2. Write down a 2×1 vector $\mathbf{u}$ with two random nonzero entries.
3. Compute $M\mathbf{u}$.

You will almost certainly find that $M\mathbf{u}$ is **not** a multiple of $\mathbf{u}$. Using the example matrix A from above as a random matrix M and the vector $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ as a random vector $\mathbf{u}$,

$$A\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 & 4 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 10 \\ 1 \end{bmatrix}.$$

There is no number λ such that $A\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \lambda\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ demonstrating that $M\mathbf{u}$ is no scalar multiple of $\mathbf{u}$. Multiplying a vector by A does not always produce a multiple of the vector.

Letting $B = \begin{bmatrix} 3 & 0 \\ -1 & -2 \end{bmatrix}$, $B\mathbf{v}_1 = \begin{bmatrix} 3 & 0 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -3 \\ -3 \end{bmatrix}$ but $\begin{bmatrix} -3 \\ -3 \end{bmatrix} \neq \lambda\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ for any value of λ. So $\mathbf{v}_1$ is not some magical vector such that multiplying it by any matrix gives a multiple. That kind of vector does not exist.

From the evidence A is not special on its own, nor are $\mathbf{v}_1$ or $\mathbf{v}_2$ special on their own. A and $\mathbf{v}_1$ are only special together just as A and $\mathbf{v}_2$ are only special together. To indicate the special relationship between A and $\mathbf{v}_1$ ($A\mathbf{v}_1 = \lambda\mathbf{v}_1$ for some λ), we call $\mathbf{v}_1$ an **eigenvector** of A. Similarly, $\mathbf{v}_2$ is an eigenvector of A. But that doesn't tell the whole story. $A\mathbf{v}_1 = \lambda\mathbf{v}_1$ and $A\mathbf{v}_2 = \lambda\mathbf{v}_2$ for different values of λ. The value -3 is associated with the eigenvector $\mathbf{v}_1$ and the value 2 is associated with the eigenvector $\mathbf{v}_2$. To mark this relationship, we call -3 an **eigenvalue** of A associated with eigenvector $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ and we call 2 an eigenvalue of A associated with the eigenvector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Any eigenvalue together with an associated eigenvector is called an **eigenpair**. For each eigenvector there is an eigenvalue, and for each eigenvalue there is an eigenvector (or is there?).

It is true, for any matrix A and any scalar λ, that $A\mathbf{0} = \lambda\mathbf{0}$ where $\mathbf{0}$ is the proper size vector with a zero for each entry, a so-called **zero vector**. Given the truth of this statement for all matrices, it does not tell us anything useful about a matrix. Moreover, if we allow $\mathbf{0}$ to be an eigenvector, the eigenvalue associated with $\mathbf{0}$ would be ill-defined. It could be any number! Therefore we disallow $\mathbf{0}$ from the definition of eigenvector. With this restriction, every eigenvector has a unique associated eigenvalue.

Given an eigenvector $\mathbf{v}$ of a matrix M, it is easy to calculate the associated eigenvalue. Given an eigenvalue of a matrix, finding an associated eigenvector takes some work. Suppose λ is an eigenvalue of the matrix M. By definition, the associated eigenvector $\mathbf{v}$ satisfies $M\mathbf{v} = \lambda\mathbf{v}$. Equivalently,

$$(M - \lambda I)\mathbf{v} = \mathbf{0}. \tag{1.7.1}$$

We will see why these equations are equivalent later on. The equivalent form (1.7.1) gives us a way to find eigenvectors. Each side of the equation expresses a vector, and for two vectors to be equal, their corresponding entries must be equal. For a vector $\mathbf{v}$ with n entries, this observation yields n linear equations in the n unknown entries of $\mathbf{v}$. Recalling how to solve linear systems of equations gives a solution for $\mathbf{v}$. To illustrate, we find an eigenvector of $\begin{bmatrix} -31 & 14 & 10 \\ -76 & 35 & 26 \\ 18 & -9 & -8 \end{bmatrix}$ associated with the eigenvalue -2. Starting with (1.7.1), we have

$$\left(\begin{bmatrix} -31 & 14 & 10 \\ -76 & 35 & 26 \\ 18 & -9 & -8 \end{bmatrix} - \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \right) \mathbf{v} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

or

$$\begin{bmatrix} -29 & 14 & 10 \\ -76 & 37 & 26 \\ 18 & -9 & -6 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Multiplying yields

$$\begin{bmatrix} -29v_1 + 14v_2 + 10v_3 \\ -76v_1 + 37v_2 + 26v_3 \\ 18v_1 - 9v_2 - 6v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

so we must have

$$\begin{array}{rcrcrcl} -29v_1 & + & 14v_2 & + & 10v_3 & = & 0 \\ -76v_1 & + & 37v_2 & + & 26v_3 & = & 0 \\ 18v_1 & - & 9v_2 & - & 6v_3 & = & 0 \end{array} \tag{1.7.2}$$

Can you find a single set of values for the variables v_1, v_2, v_3 that solves all three equations? Answer on page 46. One solution is $v_1 = v_2 = 2$ and $v_3 = 3$. We can verify that

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} \text{ is indeed an eigenvector}$$

by multiplying:

$$\begin{bmatrix} -31 & 14 & 10 \\ -76 & 35 & 26 \\ 18 & -9 & -8 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -4 \\ -4 \\ -6 \end{bmatrix} = -2 \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}.$$

Note that if $\mathbf{v}$ is an eigenvector of A associated with value λ, so is $c\mathbf{v}$. Therefore, solving for an eigenvector will always yield an infinite number of solutions. There is no unique eigenvector associated with a given eigenvalue. We will prove these facts later.

Now we can find an eigenvector of a matrix M given an eigenvalue, and we can find an eigenvalue of a matrix M given an eigenvector, but what if we have neither an eigenvalue nor eigenvector of M? Returning to (1.7.1), we know $(M - \lambda I)\mathbf{v} = \mathbf{0}$. Certainly if $\mathbf{v} = \mathbf{0}$, the equation is true. But we have decided that $\mathbf{0}$ is excluded from being an eigenvector, so we seek solutions where $\mathbf{v} \neq \mathbf{0}$. Suppose $\det(M - \lambda I) \neq 0$, meaning $(M - \lambda I)$ is invertible. Then left-multiplying both sides of (1.7.1) yields

$$(M - \lambda I)^{-1} \left((M - \lambda I)\mathbf{v} \right) = (M - \lambda I)^{-1}\mathbf{0} = \mathbf{0}.$$

But multiplication by a matrix's inverse undoes multiplication by that matrix, so $(M-\lambda I)^{-1}((M-\lambda I)\mathbf{v}) = \mathbf{v}$. Now we have that $(M-\lambda I)^{-1}((M-\lambda I)\mathbf{v}) = \mathbf{0}$ and $(M-\lambda I)^{-1}((M-\lambda I)\mathbf{v}) = \mathbf{v}$, so we conclude $\mathbf{v} = \mathbf{0}$, which is disallowed as an eigenvector. No solutions come from letting $\det(M-\lambda I) \neq 0$, so it must be that $\det(M-\lambda I) = 0$. Since the determinant of a matrix is a scalar, this equation is a scalar equation (like those you have seen in algebra) in the single variable λ. Solving that equation for λ gives eigenvalues and knowing eigenvalues gives eigenvectors.

Since

$$\det(M-\lambda I) = 0 \tag{1.7.3}$$

is the linchpin in finding eigenvalues and eigenvectors of a matrix, it has a name—the **characteristic equation** (of M). The expression $\det(M-\lambda I)$ is an n^{th} degree polynomial in λ and is called the **characteristic polynomial** (of M). Perhaps solving equations of the form *polynomial* $= 0$ reminds you of factoring. If you know that $(x-2)(x+5) = 0$, for example, then you know that either $x-2=0$ or $x+5=0$ giving two solutions, $x=2$ and $x=-5$. The equation $\det(M-\lambda I) = 0$ takes exactly this form and, when contrived as such, will be factorable.

Exercise 18 of section 3.7 requests an argument that $\det(M-\lambda I) = 0$ if and only if λ is an eigenvalue of M.

Key Concepts

characteristic equation $\det(M-\lambda I) = 0$ for any matrix M.

characteristic polynomial $\det(M-\lambda I)$ for any matrix M.

eigenpair An eigenvalue together with an associated eigenvector. $(\lambda, \mathbf{v})$ is an eigenpair for matrix M if $M\mathbf{v} = \lambda\mathbf{v}$.

eigenvalue A value λ is an eigenvalue of the matrix M if there is a nonzero vector $\mathbf{v}$ such that $M\mathbf{v} = \lambda\mathbf{v}$.

eigenvector A nonzero vector $\mathbf{v}$ is an eigenvector of a matrix M if there is a value λ such that $M\mathbf{v} = \lambda\mathbf{v}$.

zero vector Any vector with a zero for each entry.

SageMath

If `M` is a matrix in SageMath, then `M.eigenvectors_right()` lists its eigenvalues and eigenvectors, `M.eigenvalues()` lists only its eigenvalues, and `M.charpoly()` gives its characteristic polynomial. The following code computes the eigenvalues, eigenvectors, and characteristic polynomial of

$$A = \begin{bmatrix} -112 & -21 & -15 & -372 \\ -84 & -13 & -19 & -292 \\ -36 & -13 & -3 & -116 \\ 36 & 7 & 5 & 120 \end{bmatrix}.$$

```
M = matrix(4,4,[-112,-21,-15,-372,-84,-13,-19,-292,
                -36,-13,-3,-116,36,7,5,120])
print(M); print()
print(M.eigenvalues()); print()
print(M.eigenvectors_right()); print()
print(M.charpoly())
```

SageMathCell 37 The output of the code is

```
[-112  -21  -15 -372]
[ -84  -13  -19 -292]
[ -36  -13   -3 -116]
[  36    7    5  120]

[-4, -12, 4, 4]

[(-4, [
```

```
(1, -3, -3, 0)
], 1), (-12, [
(1, 2/3, 2/3, -1/3)
], 1), (4, [
(1, -1/3, 1, -1/3)
], 2)]

x^4 + 8*x^3 - 64*x^2 - 128*x + 768
```

Since the characteristic polynomial of an $n \times n$ matrix has degree n it has n eigenvalues counting multiplicities and complex eigenvalues. The output of the `eigenvalues()` method above shows that 4 has multiplicity 2. It is listed twice in the list of eigenvalues (`[-4, -12, 4, 4]`). The `eigenvectors_right()` gives the same information and more. It prints out each eigenvalue, all associated eigenvectors, and finally the multiplicity of the eigenvalue. In the output above, the `eigenvectors_right()` method shows the multiplicity of the eigenvalue 4 by listing the eigenvalue 4 followed by its associated eigenvector $\langle 1, -1/3, 1, -1/3 \rangle$ and finally its multiplicity 2 (the last 3 lines of output before the characteristic polynomial):

```
], 1), (4, [
(1, -1/3, 1, -1/3)
], 2)]
```

Exercises

1. Use the fact that $\mathbf{v}$ is an eigenvector of A to find an eigenvalue of A.

   (a) $A = \begin{bmatrix} 8 & 6 \\ -9 & -7 \end{bmatrix}$; $\mathbf{v} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$

   (b) $A = \begin{bmatrix} -5 & -4 \\ 2 & 1 \end{bmatrix}$; $\mathbf{v} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ [𝔸]-348

   (c) $A = \begin{bmatrix} -4 & 1 & 1 \\ 2 & 0 & -2 \\ -4 & -1 & 1 \end{bmatrix}$; $\mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$ [𝕊]-287

   (d) $A = \begin{bmatrix} 24 & -8 & 10 \\ 0 & 6 & 0 \\ -45 & 18 & -21 \end{bmatrix}$; $\mathbf{v} = \begin{bmatrix} 2 \\ 0 \\ -3 \end{bmatrix}$

   (e) $A = \begin{bmatrix} -28 & 0 & 48 & -48 \\ 80 & 6 & -186 & 182 \\ -4 & -1 & -3 & 9 \\ 4 & -1 & -15 & 21 \end{bmatrix}$; $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \\ 2 \\ 1 \end{bmatrix}$

   (f) $A = \begin{bmatrix} -24 & -8 & 11 & 27 \\ 216 & 52 & -130 & -258 \\ -52 & -8 & 29 & 57 \\ 52 & 8 & -33 & -61 \end{bmatrix}$; $\mathbf{v} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ -1 \end{bmatrix}$ [𝔸]-348

2. Find the characteristic polynomial.

   (a) $\begin{bmatrix} 7 & -10 \\ -11 & 5 \end{bmatrix}$

   (b) $\begin{bmatrix} -2 & -4 \\ 7 & 12 \end{bmatrix}$

   (c) $\begin{bmatrix} -12 & 12 \\ 9 & -9 \end{bmatrix}$ [𝔸]-348

   (d) $\begin{bmatrix} -8 & -3 \\ 3 & -2 \end{bmatrix}$ [𝕊]-287

   (e) $\begin{bmatrix} 1 & 0 & 2 \\ 4 & 5 & 1 \\ -6 & -6 & -4 \end{bmatrix}$ [𝔸]-348

   (f) $\begin{bmatrix} -12 & 9 & 6 \\ -15 & 12 & 6 \\ -5 & 3 & 5 \end{bmatrix}$

   (g) $\begin{bmatrix} 9 & -15 & 4 \\ 39 & -33 & 28 \\ 33 & -15 & 28 \end{bmatrix}$

3. Find the eigenvalues. They may be complex.

   (a) $\begin{bmatrix} 3 & -5 \\ 2 & -4 \end{bmatrix}$

   (b) $\begin{bmatrix} 8 & -10 \\ 5 & -7 \end{bmatrix}$ [𝔸]-348

   (c) $\begin{bmatrix} -9 & 4 \\ -36 & 15 \end{bmatrix}$

   (d) $\begin{bmatrix} -7 & 25 \\ -1 & 3 \end{bmatrix}$ [𝕊]-288

   (e) $\begin{bmatrix} 2 & -1 \\ -4 & 2 \end{bmatrix}$

   (f) $\begin{bmatrix} 9 & 6 \\ -15 & -10 \end{bmatrix}$ [𝔸]-348

   (g) $\begin{bmatrix} 5 & -2 \\ 6 & 3 \end{bmatrix}$

   (h) $\begin{bmatrix} -2 & 2 \\ -1 & -2 \end{bmatrix}$ [𝔸]-348

(i) $\begin{bmatrix} -1 & 10 & 6 \\ 2 & 3 & 2 \\ -2 & -2 & -1 \end{bmatrix}$

(j) $\begin{bmatrix} 2 & -3 & -2 \\ 12 & -17 & -12 \\ -15 & 21 & 15 \end{bmatrix}$ [𝔸]-348

(k) $\begin{bmatrix} -3 & 0 & 0 \\ 4 & 13 & -12 \\ 4 & 16 & -15 \end{bmatrix}$ [𝕊]-288

(l) $\begin{bmatrix} -4 & 3 & 3 & 0 \\ -16 & 1 & -13 & -24 \\ -4 & 2 & 4 & 0 \\ 6 & -3 & -3 & 2 \end{bmatrix}$

4. Find an eigenvector associated with the given eigenvalue.

(a) $A = \begin{bmatrix} 3 & -10 \\ 8 & -15 \end{bmatrix}; \lambda = -5$

(b) $A = \begin{bmatrix} -6 & 20 \\ -1 & 3 \end{bmatrix}; \lambda = -2$ [𝔸]-348

(c) $A = \begin{bmatrix} -4 & 2 \\ -16 & 8 \end{bmatrix}; \lambda = 0$ [𝕊]-288

(d) $A = \begin{bmatrix} 1 & 6 \\ 3 & -5 \end{bmatrix}; \lambda = -2 + 3\sqrt{3}$

(e) $A = \begin{bmatrix} 2 & 4 \\ -3 & -4 \end{bmatrix}; \lambda = -1 - i\sqrt{3}$ [𝔸]-348

(f) $A = \begin{bmatrix} -5 & 6 & -12 \\ 7 & -8 & 16 \\ 5 & -6 & 12 \end{bmatrix}; \lambda = -2$

(g) $A = \begin{bmatrix} 9 & 1 & -5 \\ 33 & 17 & -25 \\ 36 & 12 & -24 \end{bmatrix}; \lambda = 6$

(h) $A = \begin{bmatrix} -18 & 6 & 6 \\ -19 & 5 & 9 \\ -11 & 5 & 1 \end{bmatrix}; \lambda = 0$ [𝔸]-348

5. Describe a connection between eigenvalues and determinants.

6. True or false? [𝔸]-348

(a) The only eigenvector corresponding to a zero eigenvalue is the zero vector.

(b) An eigenvalue may be any complex number except zero.

(c) An eigenvector cannot be the zero vector.

(d) All 2×2 matrices have two different eigenvalues.

(e) Each eigenvalue has exactly one corresponding eigenvector.

(f) An eigenvalue may be any number, including zero and complex numbers.

(g) An $n \times n$ matrix can have $n + 1$ eigenvalues.

7. True or false? If all the entries in a square matrix M are integers but one of its eigenvalues is $\sqrt{2}$, then the entries of the corresponding eigenvector cannot all be integers.

8. Suppose matrix A is a 3×3 matrix such that $A \cdot \begin{bmatrix} 20 \\ -16 \\ 8 \end{bmatrix} = \begin{bmatrix} 5 \\ -4 \\ 2 \end{bmatrix}$. Find an eigenvalue of A.

9. SageMathCell 38 Let

$$M = \begin{bmatrix} 99 & -135 & -199 & 417 \\ 30 & -36 & -61 & 123 \\ 90 & -135 & -174 & 369 \\ 30 & -45 & -57 & 120 \end{bmatrix}$$

and use SageMath to determine which of the following vectors is an eigenvector of M. Also determine its associated eigenvalue. [𝔸]-348

$$\begin{bmatrix} 4 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -6 \\ -3 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ -6 \\ 5 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ -6 \\ -2 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 1 \\ -1 \end{bmatrix}$$

10. SageMathCell 39 Use SageMath's

```
M.eigenvectors_right()
```

method to compute the eigenvectors of M from question 9. Notice that none of the vectors from that question is listed as an eigenvector by SageMath, yet one of them is. Can you resolve this conundrum? Is it possible that linear combinations of eigenvectors are also eigenvectors?

11. SageMathCell Check your work on question 2 using SageMath's `charpoly()` method. [𝕊]-289

Answers

system solution: One possible solution of equations (1.7.2) follows. Start by dividing the third equation by 3, which yields

$$\begin{array}{rcrcrcl} -29v_1 & + & 14v_2 & + & 10v_3 & = & 0 \\ -76v_1 & + & 37v_2 & + & 26v_3 & = & 0 \\ 6v_1 & - & 3v_2 & - & 2v_3 & = & 0 \end{array}$$

The v_3 variable can be eliminated from the first two equations by adding 5 times the third equation to the first and 13 times the third equation to the second:

$$\begin{array}{rcrcrcl} v_1 & - & v_2 & & & = & 0 \\ 2v_1 & - & 2v_2 & & & = & 0 \\ 6v_1 & - & 3v_2 & - & 2v_3 & = & 0 \end{array}$$

Dividing the second equation by 2, we see it is just a repeat of the first equation, $v_1 - v_2 = 0$, which implies that $v_1 = v_2$. Substituting into the third equation, we find

$$6v_2 - 3v_2 - 2v_3 = 0$$

or $3v_2 - 2v_3 = 0$, which means $v_3 = \frac{3}{2}v_2$. This set of equations has infinitely many solutions! They take the form $v_1 = v_2$, $v_3 = \frac{3}{2}v_2$, and v_2 is any number. In terms of the eigenvector, this observation means

$$\mathbf{v} = \begin{bmatrix} v_2 \\ v_2 \\ \frac{3}{2}v_2 \end{bmatrix} = v_2 \begin{bmatrix} 1 \\ 1 \\ \frac{3}{2} \end{bmatrix}$$

for some value v_2. If $v_2 = 0$, however, then $\mathbf{v} = \mathbf{0}$, which is disallowed as an eigenvector. Therefore *every nonzero scalar multiple* of $\begin{bmatrix} 1 \\ 1 \\ \frac{3}{2} \end{bmatrix}$ is an eigenvector of A associated with the eigenvalue -2.

Chapter 2

Row Operations

2.1 Systems of Linear Equations

People have been solving systems of linear equations for millenia, since long before the advent of what we know today as algebra. Recorded history of linear systems dates back to about 150 BCE in China! Even the modern process of elimination, often learned in high school algebra and precalculus classes, and demonstrated on page 46, dates back in geometric form to at least the second century CE (circa 100), appearing as a narrative in the Chinese treatise *Nine Chapters on the Mathematical Arts*. Roger Hart proposes that a geometric version of the modern day procedure of using determinants to solve systems is also evident in *Nine Chapters* [11].

Crumpet 12: Yanghui Triangle

In China, the triangular arrangement of binomial coefficients, the first five rows of which are

$$\begin{array}{c} 1 \\ 1\ 1 \\ 1\ 2\ 1 \\ 1\ 3\ 3\ 1 \\ 1\ 4\ 6\ 4\ 1 \end{array}$$

is often called the Yanghui triangle. It was devised in the 11^{th} century CE by Jia Xian and popularized in the 13^{th} century by Yang Hui. Omar Khayyam, an 11^{th} century Persian figure also studied the triangle. [13]

In our modern treatment of systems of linear equations, a **linear equation** is an equation that takes the form

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = b \tag{2.1.1}$$

for coefficients $c_1, c_2, \ldots, c_n$, variables $v_1, v_2, \ldots, v_n$ and constant b. The (ordered) list $s_1, s_2, \ldots, s_n$ is a **solution of the equation** if and only if substitution of the values $s_1, s_2, \ldots, s_n$ for the variables $v_1, v_2, \ldots, v_n$, respectively, make the equation true. A set of linear equations is called a **linear system** or **system of linear equations**. The (ordered) list $s_1, s_2, \ldots, s_n$ is a **solution of a linear system** if and only if it is a solution of every equation in the system. The set of equations (1.7.2) on page 43 is an example of a system of linear equations, and the lists $1, 1, \frac{3}{2}$ and $2, 2, 3$ are solutions of the system (given that these lists specify values of the variables v_1, v_2, v_3 in that order).

Calculations like the one on page 46 easily submit to the conciseness of matrices. If you are familiar with synthetic division, you are already familiar with this idea. Only coefficients and constants are retained. All variables and other "extraneous" symbols are unused. On the left is the original calculation from page 46. On the right is an accounting of the coefficients and constants of each equation during each step of the process, maintained in a 3×4 matrix. The first column holds the v_1 coefficients, the second column holds the v_2 coefficients, the third column holds the v_3 coefficients, and the fourth column holds the constants from the righthand sides of the equations.

The given system:

$$\begin{array}{rcrcrcl} -29v_1 & + & 14v_2 & + & 10v_3 & = & 0 \\ -76v_1 & + & 37v_2 & + & 26v_3 & = & 0 \\ 6v_1 & - & 3v_2 & - & 2v_3 & = & 0 \end{array} \tag{2.1.2}$$

As a matrix:

$$\begin{bmatrix} -29 & 14 & 10 & 0 \\ -76 & 37 & 26 & 0 \\ 6 & -3 & -2 & 0 \end{bmatrix}$$

Adding 5 times the third equation to the first and 13 times the third equation to the second:

$$\begin{array}{rcrcrcl} v_1 & - & v_2 & & & = & 0 \\ 2v_1 & - & 2v_2 & & & = & 0 \\ 6v_1 & - & 3v_2 & - & 2v_3 & = & 0 \end{array}$$

Adding 5 times the third row to the first and 13 times the third row to the second:

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

Dividing the second equation by 2, we see it is just a repeat of the first equation, $v_1 - v_2 = 0$, so we can scrap it:

$$\begin{array}{rcrcrcl} v_1 & - & v_2 & & & = & 0 \\ 6v_1 & - & 3v_2 & - & 2v_3 & = & 0 \end{array}$$

Dividing the second row by 2, we see it is just a repeat of the first row, 1 -1 0 0, so we can zero it out and swap it with the third row:

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

Substituting $v_1 = v_2$ into the equation $6v_1 - 3v_2 - 2v_3 = 0$:

$$3v_2 - 2v_3 = 0$$

Subtracting 6 times the first row from the second:

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

In either case, we have reduced the system to the two equations $v_1 - v_2 = 0$ and $3v_2 - 2v_3 = 0$, from which the solution follows.

Much like successful synthetic division is dependent on strict ordering of the coefficients of the polynomial, it should be noted that the success of the matrix process is dependent on strict ordering of the entries of the matrix. Each row of the matrix represents one equation. The rightmost column of the matrix represents the constants from the righthand sides of the equations. Each of the remaining columns represents the coefficients of a single variable from the lefthand sides of the equations. It can easily be verified that the systems

$$\begin{array}{rcrcrcl} -29v_1 & + & 14v_2 & + & 10v_3 & = & 0 \\ -76v_1 & + & 37v_2 & + & 26v_3 & = & 0 \\ 6v_1 & - & 3v_2 & - & 2v_3 & = & 0 \end{array} \quad \text{and} \quad \begin{array}{rcl} 14v_2 + 10v_3 & = & 29v_1 \\ 26v_3 + 37v_2 & = & 76v_1 \\ 6v_1 & = & 2v_3 + 3v_2 \end{array}$$

are equivalent. Each equation of the system on the left has simply been rewritten using positive coefficients in the system on the right. However, the matrix representation of the system on the right, which might be written as

$$\begin{bmatrix} 14 & 10 & 29 & 0 \\ 26 & 37 & 76 & 0 \\ 0 & 6 & 2 & 3 \end{bmatrix}$$

is not helpful in solving the system using row operations. Certainly we can subtract 38 times the third row from the second row:

$$\begin{bmatrix} 26 & 37 & 76 & 0 \end{bmatrix} - 38 \begin{bmatrix} 0 & 6 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 26 & -191 & 0 & -114 \end{bmatrix}$$

to obtain the matrix

$$\begin{bmatrix} 14 & 10 & 29 & 0 \\ 26 & -191 & 0 & -114 \\ 0 & 6 & 2 & 3 \end{bmatrix}$$

creating a zero in the 2,3-entry just as before. The problem is the -191 and 0 of the resultant matrix have no meaning in terms of the original system. Those parts of the calculation represent $37v_2-38(6v_1)$, which simplifies to $37v_2-228v_1$ and not necessarily -191 times anything; and $76v_1-38(2v_3)$, which simplifies to $76v_1-76v_3$ or $76(v_1-v_3)$ and not necessarily 0 times anything. For row operations to make sense in the context of solving systems of equations, the entries in a single column must all be coefficients of the same variable or constants from the equations of the system. By convention, the rightmost column always holds the constants and the remainder of the columns represent the coefficients of one variable each. The order of the columns of coefficients is flexible as long as the order is known.

Besides noting that the order of the entries in the matrix is critical, what should be taken from the matrix solution of (2.1.2) is the fact that three row operations are enough to mirror the process of elimination using matrices. On the matrix side, we swapped rows, multiplied rows by a nonzero constant, and added multiples of one row to another. Everything done in solving a linear system can be modeled by one of these operations on the corresponding matrix. As such, these three operations are called **elementary row operations**. To summarize, name them, and provide shorthand notation, the elementary row operations are

Swap: swap two rows.

Shorthand for swapping rows j and k of matrix M: $M_{j,:} \leftrightarrow M_{k,:}$

Scale: multiply each entry in a row by a nonzero scalar.

Shorthand for scaling row j of matrix M by s: $sM_{j,:} \to M_{j,:}$

Replace: add a multiple of one row to another.

Shorthand for adding s times row j to row k of matrix M: $sM_{j,:} + M_{k,:} \to M_{k,:}$

Any system of linear equations can be translated into a matrix and solved using these three operations. Even systems with no solution reveal themselves as unsolvable under the direction of these three operations. If that were the only purpose of row operations, they would be useful, but as the concepts of linear algebra unfold, the ideas laid out here will have many further reaching consequences.

Elementary Matrices

Any matrix resulting from performing an elementary row operation on an identity matrix is called an **elementary matrix**. For example

$$E = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is an elementary matrix since it is just $I_{4\times 4}$ with the first two rows swapped. The feature of interest is that left-multiplying any other matrix by this elementary matrix has the effect of performing the corresponding row operation (swapping the first two rows) on that arbitrary matrix so long as the product is defined. For example, if we let $A = \begin{bmatrix} 0 & -12 & 4 & -10 & 10 \\ 6 & 2 & -8 & -2 & 5 \\ -5 & -4 & -6 & 9 & 7 \\ 1 & -1 & 8 & -9 & 3 \end{bmatrix}$, then

$$EA = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -12 & 4 & -10 & 10 \\ 6 & 2 & -8 & -2 & 5 \\ -5 & -4 & -6 & 9 & 7 \\ 1 & -1 & 8 & -9 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 2 & -8 & -2 & 5 \\ 0 & -12 & 4 & -10 & 10 \\ -5 & -4 & -6 & 9 & 7 \\ 1 & -1 & 8 & -9 & 3 \end{bmatrix}. \tag{2.1.3}$$

Take a short break to verify at least one of the entries in each row. This exercise will help you see why the product works out as shown and will help illuminate the following computations.

The $1,j$-entry of EA is a linear combination of the entries in the j^{th} column of A. To be precise, $(EA)_{1,j} = 0A_{1,j} + 1A_{2,j} + 0A_{3,j} + 0A_{4,j}$. The entries from the first row of E are used as the coefficients of the linear combination needed to compute each entry of the first row of EA. It follows that the first row of EA is a linear combination of the rows of A using these same coefficients. Symbolically, $(EA)_{1,:} = 0A_{1,:} + 1A_{2,:} + 0A_{3,:} + 0A_{4,:}$. In summary, each row of EA is the linear combination of the rows of A with coefficients from the corresponding row of E. To compute the third row of EA, for example, we use the entries from the third row of E as coefficients of a linear combination of the rows of A:

$$\begin{aligned}(EA)_{3,:} =& 0\begin{bmatrix} 0 & -12 & 4 & -10 & 10 \end{bmatrix}+\\ & 0\begin{bmatrix} 6 & 2 & -8 & -2 & 5 \end{bmatrix}+\\ & 1\begin{bmatrix} -5 & -4 & -6 & 9 & 7 \end{bmatrix}+\\ & 0\begin{bmatrix} 1 & -1 & 8 & -9 & 3 \end{bmatrix}\\ =& \begin{bmatrix} -5 & -4 & -6 & 9 & 7 \end{bmatrix}.\end{aligned}$$

All the rows of EA can be computed quickly from this perspective. The first row of E, $\begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}$, suggests that the first row of EA is 0 times row one of A plus 1 times row 2 of A plus 0 times row 3 of A plus 0 times row 4 of A. Looking at it this way makes it clear that the first row of EA is simply the second row of A. Similarly, the second row of EA is just the first row of A, the third row of EA is the third row of A, and the fourth row of EA is the fourth row of A. In other words, multiplying A by E has the effect of swapping the first two rows of A as claimed. The effect of multiplying by other elementary matrices can be verified similarly.

Key Concepts

elementary matrix The matrix resulting from performing a single row operation on an identity matrix.

elementary row operation One of swap, scale, or replace.

linear equation An equation of the form $c_1v_1+c_2v_2+\cdots+c_nv_n = b$ where $c_1, c_2, \ldots, c_n$ are coefficients, $v_1, v_2, \ldots, v_n$ are variables, and b is a constant.

linear system Another name for a system of linear equations.

row swap Swapping two rows of a matrix.

row scale Multiplying each entry of a single row of a matrix by a nonzero scalar.

row replace Replacing a row of a matrix with the sum of it plus some multiple of another row.

system of linear equations A set of linear equations.

solution of a linear system An (ordered) list of values $s_1, s_2, \ldots, s_n$ that is a solution of every equation in the system.

solution of a linear equation An (ordered) list of values $s_1, s_2, \ldots, s_n$ such that substitution for the variables $v_1, v_2, \ldots, v_n$, respectively, in the linear equation $c_1v_1 + c_2v_2 + \cdots + c_nv_n = b$ make the equation true.

Exercises

1. Represent the linear system as a matrix.

(a) $$\begin{array}{rcrcrcl} -x & - & 4y & + & 8z & = & 0 \\ 6x & + & 7y & - & 8z & = & 6 \\ -5x & - & 7y & + & 9z & = & 9 \end{array}$$

(b) $$\begin{array}{rcrcrcr} 3x & + & 2y & - & 8z & = & 9 \\ & - & 3y & + & 2z & = & 10 \\ -7x & + & y & & & = & -11 \end{array}$$ [𝔸]-348

(c) $$\begin{array}{rcrcrcl} 2x_1 & - & 8 & + & x_3 & = & 0 \\ 3 & + & x_2 & - & x_3 & = & 0 \\ x_1 & + & 2x_2 & - & 5 & = & 0 \end{array}$$

(d) $$\begin{array}{rcrcrcrcr} 3v_1 & + & 2v_2 & + & 7 & = & 3v_3 & & \\ & & 8 & + & v_2 & = & 5v_1 & + & 2v_3 \\ v_1 & + & v_2 & + & v_3 & = & 11 & & \end{array}$$ [𝔸]-348

2. Write the linear system associated with the matrix. Assume the variables of the system are $v_1, v_2, \ldots, v_n$, their coefficients appear in that order in the matrix, and the rightmost column holds the constants.

(a) $$\begin{bmatrix} 5 & -7 & 8 \\ -8 & 6 & 2 \end{bmatrix}$$

(b) $$\begin{bmatrix} -14 & -15 & 8 & -8 \\ -13 & 2 & -1 & 13 \\ 15 & -9 & -6 & 12 \end{bmatrix}$$ [𝔸]-348

(c) $\begin{bmatrix} 6 & 0 & -11 \\ 4 & 12 & -1 \\ -5 & 15 & 1 \end{bmatrix}$

(d) $\begin{bmatrix} 10 & -9 & -1 & 3 & 15 & 6 \\ -11 & 12 & 13 & 5 & -4 & -2 \end{bmatrix}$ [A]-348

3. The matrix for a linear system is given. Find one solution of the associated system. Assume the variables of the system are $v_1, v_2, \ldots v_n$ and their coefficients appear in that order in the matrix.

(a) $\begin{bmatrix} 1 & 0 & 0 & 6 \\ 0 & 3 & 0 & 8 \\ 0 & 0 & 1 & -2 \end{bmatrix}$

(b) $\begin{bmatrix} 11 & 0 & 0 & 0 & 9 \\ 0 & 5 & 0 & 0 & -7 \\ 0 & 0 & 1 & 0 & -13 \\ 0 & 0 & 0 & -2 & 6 \end{bmatrix}$ [S]-289

(c) $\begin{bmatrix} 9 & -9 & -2 \\ 0 & 3 & 4 \end{bmatrix}$

(d) $\begin{bmatrix} 11 & 0 & 9 & 12 \\ 0 & -8 & -4 & -1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$ [S]-289

4. A matrix representing a linear system is given. Explain why the system has no solution.

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}$$

5. A matrix representing a linear system is given. Find one solution of the system, and explain why the system has infinitely many more solutions. [A]-348

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

6. What elementary row operation will transform the first matrix into the second?

(a) $\begin{bmatrix} 1 & -4 & 4 & -9 \\ 1 & 6 & -5 & 1 \\ -2 & 7 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -4 & 4 & -9 \\ 1 & 6 & -5 & 1 \\ 0 & 19 & -9 & 3 \end{bmatrix}$

(b) $\begin{bmatrix} -3 & 3 & 3 & -7 \\ 4 & -3 & 3 & 6 \\ -9 & -6 & 9 & -1 \end{bmatrix}, \begin{bmatrix} -3 & 3 & 3 & -7 \\ -9 & -6 & 9 & -1 \\ 4 & -3 & 3 & 6 \end{bmatrix}$ [A]-348

(c) $\begin{bmatrix} -5 & 3 & -2 & 7 \\ 2 & -6 & 3 & -5 \\ -9 & -9 & 7 & -9 \end{bmatrix}, \begin{bmatrix} -9 & -9 & 7 & -9 \\ 2 & -6 & 3 & -5 \\ -5 & 3 & -2 & 7 \end{bmatrix}$

(d) $\begin{bmatrix} 5 & -4 & 8 & 2 \\ 9 & 3 & -2 & -2 \\ 1 & -5 & -9 & -2 \end{bmatrix}, \begin{bmatrix} 5 & -4 & 8 & 2 \\ 0 & 48 & 79 & 16 \\ 1 & -5 & -9 & -2 \end{bmatrix}$ [A]-348

(e) $\begin{bmatrix} 2 & 6 & -1 & -8 \\ -1 & 1 & -6 & 4 \\ 2 & -5 & -6 & 6 \end{bmatrix}, \begin{bmatrix} 2 & 6 & -1 & -8 \\ -2 & 2 & -12 & 8 \\ 2 & -5 & -6 & 6 \end{bmatrix}$

7. What elementary row matrix E makes the equation true?

(a) $E \cdot \begin{bmatrix} 12 & -13 & 16 & -1 \\ 15 & -8 & 2 & -7 \\ -5 & -4 & -3 & -1 \end{bmatrix} = \begin{bmatrix} 12 & -13 & 16 & -1 \\ -60 & 32 & -8 & 28 \\ -5 & -4 & -3 & -1 \end{bmatrix}$

(b) $E \cdot \begin{bmatrix} 15 & 6 & 19 \\ -5 & -18 & -1 \\ 4 & 2 & 5 \end{bmatrix} = \begin{bmatrix} 15 & 6 & 19 \\ 4 & 2 & 5 \\ -5 & -18 & -1 \end{bmatrix}$

(c) $E \cdot \begin{bmatrix} 12 & 18 & -2 \\ -10 & 3 & 6 \\ 15 & 5 & 3 \end{bmatrix} = \begin{bmatrix} 12 & 18 & -2 \\ -10 & 3 & 6 \\ -30 & -10 & -6 \end{bmatrix}$

(d) $E \cdot \begin{bmatrix} 4 & -5 \\ 2 & 9 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 4 & -5 \\ -18 & -16 \\ 4 & 5 \end{bmatrix}$

(e) $E \cdot \begin{bmatrix} -4 & -14 \\ -19 & 7 \\ -20 & 17 \\ 13 & 14 \end{bmatrix} = \begin{bmatrix} 13 & 14 \\ -19 & 7 \\ -20 & 17 \\ -4 & -14 \end{bmatrix}$

(f) $E \cdot \begin{bmatrix} 8 & 9 & -7 \\ 5 & 9 & 10 \\ 9 & 3 & -4 \\ -4 & -1 & 4 \end{bmatrix} = \begin{bmatrix} 23 & 36 & 23 \\ 5 & 9 & 10 \\ 9 & 3 & -4 \\ -4 & -1 & 4 \end{bmatrix}$

8. Perform the row operation on the matrix.

(a) $A = \begin{bmatrix} 3 & -8 & -9 \\ -9 & 5 & 2 \\ 3 & -7 & -5 \\ -1 & 1 & -4 \end{bmatrix}; -2A_{3,:} \to A_{3,:}$

(b) $B = \begin{bmatrix} 3 & -1 & 1 & -7 \\ 1 & -2 & 6 & -4 \end{bmatrix}; 5B_{2,:} \to B_{2,:}$ [A]-348

(c) $C = \begin{bmatrix} -6 & 2 \\ 9 & 9 \\ 9 & 7 \\ 1 & -9 \end{bmatrix}; C_{1,:} \leftrightarrow C_{2,:}$

(d) $D = \begin{bmatrix} -9 & -9 & -6 \\ 5 & 6 & -2 \\ -9 & 9 & -4 \end{bmatrix}; D_{1,:} \leftrightarrow D_{3,:}$ [A]-348

(e) $E = \begin{bmatrix} -8 & 5 \\ -7 & -9 \\ -5 & -7 \end{bmatrix}; 3E_{2,:} + E_{3,:} \to E_{3,:}$

(f) $F = \begin{bmatrix} -5 & 1 & 9 & -6 \\ -3 & -5 & 1 & -1 \\ -8 & 3 & 2 & 2 \end{bmatrix}; -5F_{1,:} + F_{2,:} \to F_{2,:}$ [A]-348

9. The matrix T is given. What elementary row operation does left-multiplication by T perform?

(a) $\begin{bmatrix} -2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$ [S]-289

(d) $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ [𝔸]-348

(e) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ [𝔸]-348

(f) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -5 & 0 & 1 \end{bmatrix}$

10. Carry out a single elementary row operation to secure a 1 in row 1, column 1.

(a) $\begin{bmatrix} 10 & 1 & 7 \\ 3 & -3 & -5 \end{bmatrix}$

(b) $\begin{bmatrix} 5 & 6 & -4 \\ 2 & 3 & -2 \end{bmatrix}$ [𝔸]-349

(c) $\begin{bmatrix} -1 & 2 & 3 \\ -2 & 10 & -3 \end{bmatrix}$

11. Execute a single elementary row operation to secure a 0 in row 2, column 1.

(a) $\begin{bmatrix} 1 & 10 & 2 \\ 3 & -3 & -5 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 & -3 \\ 2 & 3 & -2 \end{bmatrix}$ [𝔸]-349

(c) $\begin{bmatrix} 1 & -2 & -3 \\ -2 & 10 & -3 \end{bmatrix}$

12. Find the first row of the product by computing an appropriate linear combination of the rows of the righthand matrix.

(a) $\begin{bmatrix} 7 & -8 \\ -1 & -12 \end{bmatrix}\begin{bmatrix} -3 & 0 \\ -1 & 3 \end{bmatrix}$

(b) $\begin{bmatrix} -2 & 1 \\ 7 & -5 \end{bmatrix}\begin{bmatrix} 4 & -3 & 2 \\ 12 & 8 & -5 \end{bmatrix}$ [𝕊]-290

(c) $\begin{bmatrix} -5 & -2 \\ 0 & -7 \\ 1 & 9 \end{bmatrix}\begin{bmatrix} -3 & 0 & -1 \\ 6 & 1 & 2 \end{bmatrix}$

(d) $\begin{bmatrix} 5 & -2 & 4 \\ 7 & -5 & -8 \end{bmatrix}\begin{bmatrix} 0 & 3 & 2 \\ -2 & 1 & 0 \\ 2 & 0 & 3 \end{bmatrix}$

13. If the third row of A contains all zeros, what can you say about the third row of AB? Assume the product AB is defined.

14. If the first and second rows of A are multiples of one another, what can you say about the first and second rows of AB? Assume the product AB is defined. [𝔸]-349

2.2 Row Reduction

The system associated with a matrix of the form

$$\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 5 \end{bmatrix}$$

has only one solution: $4, 5$, and we can see that without doing any computation. The system associated with this matrix, something you can probably visualize mentally, looks like

$$\begin{aligned} x &= 4 \\ y &= 5 \end{aligned}$$

(though you may have imagined different variables). The system is solved! The system corresponding to the matrix

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

corresponds to a system that is not solved. Can you write down the associated system to verify? Answer on page 61. **Row reduction** is the process of using elementary row operations to transform a matrix whose associated linear system is not solved into one whose associated linear system is solved, thus solving the system.

The following algorithm describes the process of row reduction for solving a system of n equations in n unknowns, starting with a matrix representation of the system.

Step 1: Select the leftmost column with at least one nonzero entry. This is a **pivot column**. The topmost position (row and column) in this column is a **pivot position**. If no such column exists or there are no rows below the pivot position, continue to step 5.

Step 2: If the entry in the pivot position is 0, **swap** rows so the entry in the pivot position is nonzero. This nonzero entry is a **pivot**.

Step 3: **Replace** rows until all entries in the pivot column below the pivot are 0.

Step 4: Take the submatrix consisting of all rows below the pivot and return to step 1.

Step 5: Select the column of the rightmost pivot position and **replace** rows until all entries in that column other than the pivot are 0.

Step 6: **Scale** the row with the pivot so the pivot in that row is 1.

Step 7: If there are no rows above the pivot, stop. Otherwise, take the submatrix consisting of all rows above the pivot and return to step 5.

Translating the resulting matrix back into a system of equations reveals the solution. Any matrix produced by the completion of steps 1 through 4 is said to be in **row echelon form**. The system corresponding to a matrix in row echelon form can be solved by back-substitution, so these steps of the algorithm are often sufficient. Any matrix produced by the completion of the entire algorithm is said to be in **reduced row echelon form**. The system corresponding to a matrix in reduced row echelon form is solved.

Even though the algorithm may sound rather rigid, there are choices to be made along the way. All choices will lead to the same solution, but different choices may lead to drastically different-looking routes. In the end, all choices in row reduction by hand are a matter of preference.

Crumpet 13: Automated Row Reduction

A computer programmed to perform row reduction will have to make choices just as a human working by hand would. However, the objectives of a computer are slightly different from the objectives of a human. A human is looking to make the computation as easy as possible while a computer should be programmed to make the computation as *accurate* as possible. For a computer, working with fractions or decimal values is just as easy as working with integers, and doing a couple "extra" row operations is usually preferable to doing a lengthy analysis of how to avoid them. But computer computations are subject to round-off error, something that should be minimized whenever possible. Swapping rows so the pivot is the entry of greatest magnitude in the column helps reduce round-off error. The following is essentially computer pseudo-code that reduces a matrix to row echelon form.

Step 1: Translate the system into a matrix A.

Step 2: Let $i = k = 1$.

Step 3: **Swap** row i with row j where $j > i$ and $|A_{j,k}| > |A_{m,k}|$ for all $m \geq i$. If no such row exists, increment k by one and try again as long as $k \leq n$. If k reaches $n + 1$, stop.

Step 4: **Scale** row i by $\frac{1}{A_{i,k}}$.

Step 5: For each j from $i + 1$ through n, **replace** row j by $-\frac{A_{j,k}}{Ai,k}$ times row i plus row j.

Step 6: Increment i and k each by one and return to step 3 as long as $k \leq n$.

The resulting matrix must be returned to the form of a linear system and solved by back-substitution to complete the solution.

To illustrate some of the choices that must be made, consider solving the system

$$\begin{array}{rcrcrcr} 3x_1 & - & x_2 & - & 2x_3 & = & 5 \\ 2x_1 & + & 4x_2 & + & 8x_3 & = & -13 \\ x_1 & + & 2x_2 & + & 3x_3 & = & -4 \end{array}$$

by row reduction. The following discussion charts the progress of three different approaches—(1) using fractions, (2) avoiding fractions by scaling non-pivot rows, and (3) avoiding fractions by scaling and swapping. All three approaches start from the matrix

$$A = \begin{bmatrix} 3 & -1 & -2 & 5 \\ 2 & 4 & 8 & -13 \\ 1 & 2 & 3 & -4 \end{bmatrix}$$

of the system.

(1) Using fractions	(2) Avoiding fractions by scaling	(3) Avoiding fractions by scaling and swapping
Step 1: The first column is a pivot column. The pivot position is the topmost entry of the pivot column. There is nothing to do besides note this fact. **Step 2**: The pivot position must not contain a zero. This is already the case in all three approaches, but in approach (3) the first and third rows are swapped to secure a 1 as a pivot.		
		$A_{1,:} \leftrightarrow A_{3,:}$ $$\begin{bmatrix} 1 & 2 & 3 & -4 \\ 2 & 4 & 8 & -13 \\ 3 & -1 & -2 & 5 \end{bmatrix}$$
Step 3: Replace rows to secure zeros in all rows below the pivot. Multiples of the row with the pivot are added to the rows below. In approach (1) the first row is scaled by $\frac{1}{3}$, and in approach (2) the second and third rows are scaled by 3 to prepare for replacement.		
$\frac{1}{3}A_{1,:} \to A_{1,:}$ $$\begin{bmatrix} 1 & -\frac{1}{3} & -\frac{2}{3} & \frac{5}{3} \\ 2 & 4 & 8 & -13 \\ 1 & 2 & 3 & -4 \end{bmatrix}$$	$3A_{2,:} \to A_{2,:}$ and $3A_{3,:} \to A_{3,:}$ $$\begin{bmatrix} 3 & -1 & -2 & 5 \\ 6 & 12 & 24 & -39 \\ 3 & 6 & 9 & -12 \end{bmatrix}$$	

<table>
<tr>
<td>$-2A_{1,:}+A_{2,:}\to A_{2,:}$ and
$-A_{1,:}+A_{3,:}\to A_{3,:}$
$$\begin{bmatrix} 1 & -\frac{1}{3} & -\frac{2}{3} & \frac{5}{3} \\ 0 & \frac{14}{3} & \frac{28}{3} & -\frac{49}{3} \\ 0 & \frac{7}{3} & \frac{11}{3} & -\frac{17}{3} \end{bmatrix}$$</td>
<td>$-2A_{1,:}+A_{2,:}\to A_{2,:}$ and
$-A_{1,:}+A_{3,:}\to A_{3,:}$
$$\begin{bmatrix} 3 & -1 & -2 & 5 \\ 0 & 14 & 28 & -49 \\ 0 & 7 & 11 & -17 \end{bmatrix}$$</td>
<td>$-2A_{1,:}+A_{2,:}\to A_{2,:}$ and
$-3A_{1,:}+A_{3,:}\to A_{3,:}$
$$\begin{bmatrix} 1 & 2 & 3 & -4 \\ 0 & 0 & 2 & -5 \\ 0 & -7 & -11 & 17 \end{bmatrix}$$</td>
</tr>
<tr>
<td colspan="3">Step 4: Take the submatrix consisting of all rows below the pivot and return to step 1. The first row of the matrix is fixed until step 5.</td>
</tr>
<tr>
<td>$$\begin{bmatrix} 0 & \frac{14}{3} & \frac{28}{3} & -\frac{49}{3} \\ 0 & \frac{7}{3} & \frac{11}{3} & -\frac{17}{3} \end{bmatrix}$$</td>
<td>$$\begin{bmatrix} 0 & 14 & 28 & -49 \\ 0 & 7 & 11 & -17 \end{bmatrix}$$</td>
<td>$$\begin{bmatrix} 0 & 0 & 2 & -5 \\ 0 & -7 & -11 & 17 \end{bmatrix}$$</td>
</tr>
<tr>
<td colspan="3">Step 1: The second column is a pivot column. The pivot position is the topmost entry of the pivot column. There is nothing to do besides note this fact. Step 2: The pivot position must not contain a zero. This is already the case in approaches (1) and (2), but in approach (3) the rows must be swapped.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$A_{1,:}\leftrightarrow A_{3,:}$
$$\begin{bmatrix} 0 & -7 & -11 & 17 \\ 0 & 0 & 2 & -5 \end{bmatrix}$$</td>
</tr>
<tr>
<td colspan="3">Step 3: Replace rows to secure zeros in all rows below the pivot. Multiples of the row with the pivot are added to the rows below. In approach (1) the first row is scaled by $\frac{1}{2}$; in approach (2) the second row is scaled by 2; and in approach (3) this step is already done.</td>
</tr>
<tr>
<td>$\frac{1}{2}A_{1,:}\to A_1$
$$\begin{bmatrix} 0 & \frac{7}{3} & \frac{14}{3} & -\frac{49}{6} \\ 0 & \frac{7}{3} & \frac{11}{3} & -\frac{17}{3} \end{bmatrix}$$</td>
<td>$2A_{2,:}\to A_2$
$$\begin{bmatrix} 0 & 14 & 28 & -49 \\ 0 & 14 & 22 & -34 \end{bmatrix}$$</td>
<td></td>
</tr>
<tr>
<td>$-A_{1,:}+A_{2,:}\to A_{2,:}$
$$\begin{bmatrix} 0 & \frac{7}{3} & \frac{14}{3} & -\frac{49}{6} \\ 0 & 0 & -1 & \frac{5}{2} \end{bmatrix}$$</td>
<td>$-A_{1,:}+A_{2,:}\to A_{2,:}$
$$\begin{bmatrix} 0 & 14 & 28 & -49 \\ 0 & 0 & -6 & 15 \end{bmatrix}$$</td>
<td></td>
</tr>
<tr>
<td colspan="3">Step 4: Take the submatrix consisting of all rows below the pivot and return to step 1. The first row of the submatrix (second row of the original matrix) is fixed until step 5.</td>
</tr>
<tr>
<td>$$\begin{bmatrix} 0 & 0 & -1 & \frac{5}{2} \end{bmatrix}$$</td>
<td>$$\begin{bmatrix} 0 & 0 & -6 & 15 \end{bmatrix}$$</td>
<td>$$\begin{bmatrix} 0 & 0 & 2 & -5 \end{bmatrix}$$</td>
</tr>
<tr>
<td colspan="3">Step 1: The third column is a pivot column. The pivot position is the topmost entry of the pivot column. There is nothing to do besides note this fact. Since there are no entries below the pivot, we move to step 5. At this point, our matrices are as follows. These matrices are all in row echelon form.</td>
</tr>
<tr>
<td>$$\begin{bmatrix} 1 & -\frac{1}{3} & -\frac{2}{3} & \frac{5}{3} \\ 0 & \frac{7}{3} & \frac{14}{3} & -\frac{49}{6} \\ 0 & 0 & -1 & \frac{5}{2} \end{bmatrix}$$</td>
<td>$$\begin{bmatrix} 3 & -1 & -2 & 5 \\ 0 & 14 & 28 & -49 \\ 0 & 0 & -6 & 15 \end{bmatrix}$$</td>
<td>$$\begin{bmatrix} 1 & 2 & 3 & -4 \\ 0 & -7 & -11 & 17 \\ 0 & 0 & 2 & -5 \end{bmatrix}$$</td>
</tr>
<tr>
<td colspan="3">Step 5: Select the column of the rightmost pivot position and replace rows until all entries in that column other than the pivot are 0. The pivots are $A_{1,1}$, $A_{2,2}$, and $A_{3,3}$, so the pivot postions are in columns 1, 2, and 3, and we select the third column. To prepare for the row replacements in approach (1) there is nothing to do; in approach (2) the third row is scaled by $\frac{1}{3}$; and in approach (3) the first and second rows are scaled by 2.</td>
</tr>
</table>

	$\frac{1}{3}A_{3,:} \to A_{3,:}$ $\begin{bmatrix} 3 & -1 & -2 & 5 \\ 0 & 14 & 28 & -49 \\ 0 & 0 & -2 & 5 \end{bmatrix}$	$2A_{1,:} \to A_{1,:}$ and $2A_{2,:} \to A_{2,:}$ $\begin{bmatrix} 2 & 4 & 6 & -8 \\ 0 & -14 & -22 & 34 \\ 0 & 0 & 2 & -5 \end{bmatrix}$
$\frac{14}{3}A_{3,:} + A_{2,:} \to A_{2,:}$ and $-\frac{2}{3}A_{3,:} + A_{1,:} \to A_{1,:}$ $\begin{bmatrix} 1 & -\frac{1}{3} & 0 & 0 \\ 0 & \frac{7}{3} & 0 & \frac{7}{2} \\ 0 & 0 & -1 & \frac{5}{2} \end{bmatrix}$	$14A_{3,:} + A_{2,:} \to A_{2,:}$ and $-A_{3,:} + A_{1,:} \to A_{1,:}$ $\begin{bmatrix} 3 & -1 & 0 & 0 \\ 0 & 14 & 0 & 21 \\ 0 & 0 & -2 & 5 \end{bmatrix}$	$11A_{3,:} + A_{2,:} \to A_{2,:}$ and $-3A_{3,:} + A_{1,:} \to A_{1,:}$ $\begin{bmatrix} 2 & 4 & 0 & 7 \\ 0 & -14 & 0 & -21 \\ 0 & 0 & 2 & -5 \end{bmatrix}$
Step 6: Scale the row with the pivot so the pivot in that row is 1. Fractions at this point are unavoidable.		
$-A_{3,:} \to A_{3,:}$ $\begin{bmatrix} 1 & -\frac{1}{3} & 0 & 0 \\ 0 & \frac{7}{3} & 0 & \frac{7}{2} \\ 0 & 0 & 1 & -\frac{5}{2} \end{bmatrix}$	$-\frac{1}{2}A_{3,:} \to A_{3,:}$ $\begin{bmatrix} 3 & -1 & 0 & 0 \\ 0 & 14 & 0 & 21 \\ 0 & 0 & 1 & -\frac{5}{2} \end{bmatrix}$	$\frac{1}{2}A_{3,:} \to A_{3,:}$ $\begin{bmatrix} 2 & 4 & 0 & 7 \\ 0 & -14 & 0 & -21 \\ 0 & 0 & 1 & -\frac{5}{2} \end{bmatrix}$
Step 7: Take the submatrix consisting of all rows above the pivot and return to step 5. The third row of the matrix is fixed.		
$\begin{bmatrix} 1 & -\frac{1}{3} & 0 & 0 \\ 0 & \frac{7}{3} & 0 & \frac{7}{2} \end{bmatrix}$	$\begin{bmatrix} 3 & -1 & 0 & 0 \\ 0 & 14 & 0 & 21 \end{bmatrix}$	$\begin{bmatrix} 2 & 4 & 0 & 7 \\ 0 & -14 & 0 & -21 \end{bmatrix}$
Step 5: The rightmost pivot postion is in column 2. To prepare for the row replacements in approach (1) there is nothing to do; in approach (2) the second row is scaled by $\frac{1}{7}$ while the first row is scaled by 2; and in approach (3) the second row is scaled by $\frac{1}{7}$.		
	$\frac{1}{7}A_{2,:} \to A_{2,:}$ and $2A_{1,:} \to A_{1,:}$ $\begin{bmatrix} 6 & -2 & 0 & 0 \\ 0 & 2 & 0 & 3 \end{bmatrix}$	$\frac{1}{7}A_{2,:} \to A_{2,:}$ $\begin{bmatrix} 2 & 4 & 0 & 7 \\ 0 & -2 & 0 & -3 \end{bmatrix}$
$\frac{1}{7}A_{2,:} + A_{1,:} \to A_{1,:}$ $\begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \\ 0 & \frac{7}{3} & 0 & \frac{7}{2} \end{bmatrix}$	$A_{2,:} + A_{1,:} \to A_{1,:}$ $\begin{bmatrix} 6 & 0 & 0 & 3 \\ 0 & 2 & 0 & 3 \end{bmatrix}$	$2A_{2,:} + A_{1,:} \to A_{1,:}$ $\begin{bmatrix} 2 & 0 & 0 & 1 \\ 0 & -2 & 0 & -3 \end{bmatrix}$
Step 6: Scale the row with the pivot so the pivot in that row is 1. Fractions at this point are unavoidable.		
$\frac{3}{7}A_{2,:} \to A_{2,:}$ $\begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & \frac{3}{2} \end{bmatrix}$	$\frac{1}{2}A_{2,:} \to A_{2,:}$ $\begin{bmatrix} 6 & 0 & 0 & 3 \\ 0 & 1 & 0 & \frac{3}{2} \end{bmatrix}$	$-\frac{1}{2}A_{2,:} \to A_{2,:}$ $\begin{bmatrix} 2 & 0 & 0 & 1 \\ 0 & 1 & 0 & \frac{3}{2} \end{bmatrix}$
Step 7: Take the submatrix consisting of all rows above the pivot and return to step 5. The bottom row of the submatrix (second row of the original matrix) is fixed.		
$\begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} 6 & 0 & 0 & 3 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 & 0 & 1 \end{bmatrix}$

<table>
<tr><td colspan="3">Step 5: The rightmost pivot postion is in column 1. There are no entries other than the pivot in that column, so there are no row replacements to do. Step 6: Scale the row with the pivot so the pivot in that row is 1. Fractions at this point are unavoidable.</td></tr>
<tr><td></td><td>$\frac{1}{6}A_{1,:} \to A_{1,:}$
$$\begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$</td><td>$\frac{1}{2}A_{1,:} \to A_{1,:}$
$$\begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$</td></tr>
<tr><td colspan="3">Step 7: There are no rows above the pivot, so we are done. At this point, our matrices are as follows. These matrices are all in reduced row echelon form.</td></tr>
<tr><td>$$\begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & \frac{3}{2} \\ 0 & 0 & 1 & -\frac{5}{2} \end{bmatrix}$$</td><td>$$\begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & \frac{3}{2} \\ 0 & 0 & 1 & -\frac{5}{2} \end{bmatrix}$$</td><td>$$\begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & \frac{3}{2} \\ 0 & 0 & 1 & -\frac{5}{2} \end{bmatrix}$$</td></tr>
</table>

Notice that all three approaches produced the same reduced row echelon form. This is not an accident and this point will be picked up again soon. Writing the linear system corresponding to this reduced row echelon form, we see

$$x_1 = \frac{1}{2}; \quad x_2 = \frac{3}{2}; \quad x_3 = -\frac{5}{2}.$$

The original system has this one solution.

A matrix containing the coefficients and constants of a linear system (one row for each equation, one column for the coefficients of each variable, and the rightmost column for the constants, as usual) is called an **augmented matrix**. A matrix containing the coefficients of a linear system (one row for each equation and one column for the coefficients of each variable, as usual) but not the constants is called a **coefficient matrix**. The coefficient matrix for any linear system is a submatrix of the corresponding augmented matrix. The augmented matrix simply has one more column—the column of constants for each equation.

The system

$$\begin{array}{rcrcr} 3x & - & 2y & = & 0 \\ 5x & - & 3y & = & 0 \end{array}$$

has augmented matrix

$$\begin{bmatrix} 3 & -2 & 0 \\ 5 & -3 & 0 \end{bmatrix}$$

and coefficient matrix

$$\begin{bmatrix} 3 & -2 \\ 5 & -3 \end{bmatrix}.$$

When the constants of a linear system are all zero, it is not necessary to represent the system as an augmented matrix. The coefficient matrix will do. After all, a column of zeros at the beginning of the row reduction process will be a column of zeros at the end of the process. Row operations do not change the entries of a column of zeros. For example, after swapping two rows with zeros in their j^{th} columns, the j^{th} column still has zeros in those rows. All that happened in column j was two zeros swapped places. The rest of the rows are not involved in the swap, so if their j^{th} columns held zeros before, they hold zeros after as well. Can you explain similarly why row replacement and row scaling also leave a column of zeros unchanged? Answer on page 61. A linear system whose constants are all zero is called a **homogeneous system**. Otherwise it is called a **nonhomogeneous system**.

Crumpet 14: Homogeneous Linear Differential Equations

A linear differential equation is homogeneous if its constant term is zero and nonhomogeneous otherwise.

Key Concepts

augmented matrix A matrix holding a particular arrangement of all the coefficients and constants of a linear system.

coefficient matrix A matrix holding a particular arrangement of all the coefficients but none of the constants of a linear system.

homogeneous system A linear system with constants all equal to zero.

nonhomogeneous system A linear system with at least one nonzero constant.

pivot the leading entry of a matrix in echelon form

pivot column a column containing a pivot

pivot position the row and column of a pivot

row reduction The process of using elementary row operations to transform a matrix whose associated linear system is not solved into one whose associated linear system is either solved or could be solved by back-substitution.

row echelon form An arrangement of the entries of a matrix following completion of the first four steps of the row reduction algorithm.

reduced row echelon form An arrangement of the entries of a matrix following completion of the row reduction algorithm.

Exercises

1. Reduce the matrix to row echelon form.

   (a) $\begin{bmatrix} -2 & -4 & -10 \\ -5 & 2 & -1 \end{bmatrix}$ [𝕊]-290

   (b) $\begin{bmatrix} 1 & -4 & 0 \\ 3 & 5 & 4 \end{bmatrix}$

   (c) $\begin{bmatrix} 6 & -1 & -4 \\ 4 & -3 & 1 \\ 2 & -5 & -1 \end{bmatrix}$

   (d) $\begin{bmatrix} -1 & 0 & 2 \\ 4 & 1 & -3 \\ 1 & 1 & 3 \end{bmatrix}$ [𝕊]-290

2. Reduce the matrix in question 1 to reduced row echelon form. [𝕊]-290

3. The coefficient matrix for a homogeneous linear system is given. Find one nontrivial solution (not all variables equal to zero) of the associated system if possible. Assume the variables of the system are $x_1, x_2, \ldots x_n$ and their coefficients appear in that order in the matrix.

   (a) $\begin{bmatrix} 3 & -1 \\ 0 & 5 \end{bmatrix}$ [𝕊]-290

   (b) $\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$

   (c) $\begin{bmatrix} -1 & 3 \\ 0 & 0 \end{bmatrix}$

   (d) $\begin{bmatrix} 2 & -2 & 5 \\ 0 & 7 & -1 \\ 0 & 0 & 3 \end{bmatrix}$

   (e) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

   (f) $\begin{bmatrix} 5 & 0 & -3 \\ 0 & -4 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ [𝕊]-290

   (g) $\begin{bmatrix} 5 & 3 & 0 & 0 \\ 0 & 0 & 3 & -4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ [𝔸]-349

4. Solve the nonhomogeneous system by row reduction.

   (a) $\begin{array}{rcrcr} -7x_1 & - & 2x_2 & = & -6 \\ 4x_1 & + & x_2 & = & -1 \end{array}$

   (b) $\begin{array}{rcrcr} v_1 & + & 4v_2 & = & -4 \\ -3v_1 & - & 11v_2 & = & -5 \end{array}$ [𝕊]-290

   (c) $\begin{array}{rcrcr} 2x & + & 15y & = & 6 \\ -8x & - & 65y & = & 4 \end{array}$

   (d) $\begin{array}{rcrcr} -3x & + & 5y & = & -3 \\ -15x & + & 20y & = & 6 \end{array}$ [𝔸]-349

   (e) $\begin{array}{rcrcr} x_1 & - & x_2 & = & 3 \\ -4x_1 & + & 5x_2 & = & -3 \end{array}$

   (f) $\begin{array}{rcrcr} -45x & - & 8y & = & -6 \\ -10x & - & 2y & = & 4 \end{array}$ [𝔸]-349

   (g) $\begin{array}{rcrcrcr} x & + & 3y & - & z & = & -3 \\ x & + & y & + & 2z & = & -5 \\ -3x & - & 10y & + & 4z & = & 6 \end{array}$

(h) $$\begin{array}{rcrcrcr} -3v_1 & - & 35v_2 & + & 10v_3 & = & 2 \\ 9v_1 & + & 130v_2 & - & 40v_3 & = & 2 \\ 9v_1 & + & 120v_2 & - & 35v_3 & = & -4 \end{array}$$
[$\mathbb{S}$]-291

(i) $$\begin{array}{rcrcrcr} -v_1 & + & 2v_2 & + & 8v_3 & = & -6 \\ -v_1 & + & v_2 & + & 5v_3 & = & -6 \\ -2v_1 & + & 2v_2 & + & 11v_3 & = & -1 \end{array}$$

(j) $$\begin{array}{rcrcrcr} 12x_1 & - & 9x_2 & - & 14x_3 & = & 3 \\ 24x_1 & - & 15x_2 & - & 21x_3 & = & -3 \\ -9x_1 & + & 6x_2 & + & 7x_3 & = & 2 \end{array}$$
[$\mathbb{A}$]-349

(k) $$\begin{array}{rcrcrcrcr} 2w & - & x & + & y & + & 4z & = & -3 \\ -w & + & x & & & - & 3z & = & 5 \\ -w & + & x & + & y & - & 2z & = & -2 \\ & & & - & y & - & z & = & -6 \end{array}$$

(l) $$\begin{array}{rcrcrcrcr} x_1 & - & x_2 & + & 4v_3 & - & v_4 & = & 1 \\ -2x_1 & + & 6x_2 & - & 10x_3 & + & 5x_4 & = & -3 \\ -x_1 & & & + & 2x_3 & - & 2x_4 & = & 2 \\ 3x_1 & - & 2x_2 & + & 11x_3 & - & 2x_4 & = & 6 \end{array}$$
[$\mathbb{A}$]-349

5. Let $\begin{bmatrix} -2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ be the coefficient matrix for a linear system. What can you say about the solution(s) of the system if augmented by the given column?

   (a) $\begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}$

   (b) $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

   (c) $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$; a, b, c are arbitrary real numbers

6. Let $A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -4 \end{bmatrix}$. What can you say about the solutions of the associated system if

   (a) the system is homogeneous and A is the coefficient matrix?

   (b) the system is nonhomogeneous and A is the coefficient matrix?

   (c) A is the augmented matrix?

Answers

associated system:

$$\begin{array}{rcrcr} 3x & - & 2y & = & 2 \\ 5x & - & 3y & = & 5 \end{array}$$

column of zeros: **(row scale)** After scaling a row with a zero in its j^{th} column, the j^{th} column still has a zero in that row since 0 times anything is 0. The rest of the rows are not involved in the scale, so if their j^{th} columns held zeros before, they hold zeros after as well. **(row replace)** After adding a multiple of a row with a zero in its j^{th} column, to a different row with a zero in its j^{th} column, both rows still have zeros in their j^{th} columns. 0 times anything is 0, so a 0 was added to the 0 in the row being replaced. Since 0 plus 0 is 0, the j^{th} column of that row is still zero. The rest of the rows are not involved in the replacement, so if their j^{th} columns held zeros before, they hold zeros after as well.

2.3 Existence, Uniqueness, and Echelon Forms

As augmented matrices,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad \text{and} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

represent fundamentally different linear systems. The first matrix represents the system

$$\begin{aligned} x &= 0 \\ 2y &= 3 \end{aligned}$$

which has exactly one solution: $x = 0$, $y = \frac{3}{2}$. The second matrix represents the system

$$\begin{aligned} x + 2y &= 0 \\ 0 &= 3 \end{aligned}$$

which has no solution. Even though the first equation has many solutions, the second equation will not be true for any values of x and y since 0 simply does not equal 3. The third matrix represents the system

$$\begin{aligned} x + 2y &= 3 \\ 0 &= 0 \end{aligned}$$

which has infinitely many solutions, $x = 3 - 2y$ with y arbitrary. One example is $y = 1$, $x = 1$.

The three associated linear systems have different types of solution sets. The first set of solutions contains exactly one element. The second set of solutions is empty. The third set of solutions contains infinitely many elements.

The three matrices

$$\begin{bmatrix} 7 & 10 & -3 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 7 & 0 & 0 \\ 0 & 10 & -3 \end{bmatrix}, \quad \text{and} \begin{bmatrix} 7 & 10 & 0 \\ 0 & 0 & -3 \end{bmatrix}$$

are similar to the first three in this way. One of them has an associated linear system with no solution, one with infinitely many solutions, and one with exactly one solution. Can you tell which is which? Answer on page 69.

Comparing the first set of three matrices to the second set of three matrices, you may notice that all six matrices have three 0 entries and three nonzero entries, one of each per column. Further, there are only three arrangements of the zeros. Using # to represent the nonzero numbers, the arrangements are

$$\begin{bmatrix} \# & \# & \# \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \# & \# & 0 \\ 0 & 0 & \# \end{bmatrix} \quad \text{and} \begin{bmatrix} \# & 0 & 0 \\ 0 & \# & \# \end{bmatrix}$$

corresponding to infinitely many solutions, no solution, and one solution. You can verify that each one is in row echelon form by applying the first 4 steps of the row reduction algorithm. The nonzero numbers may change, but the form (arrangement of zeros and nonzeros) will not! Additionally, no matter what nonzero numbers take the place of the #s, the number of solutions of the associated systems will not change. This is the real power of row echelon form matrices.

More generally, some of the entries of these forms could be replaced by other symbols without disrupting row echelon form. Using $\star$ to represent *any number* (including zero), matrices with the following arrangements of entries are all in row echelon form

$$\begin{bmatrix} \# & \star & \star \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \# & \star & \star \\ 0 & 0 & \# \end{bmatrix}, \quad \text{and} \begin{bmatrix} \# & \star & \star \\ 0 & \# & \star \end{bmatrix}. \tag{2.3.1}$$

These more general forms also have infinitely many, zero, and one solution(s), respectively. Imagining the associated linear systems will help you verify the number of solutions, but how can you tell they are in row echelon form?

Step 1 of the row reduction algorithm is applied to the entire matrix and then to each submatrix containing all the rows below the last identified pivot. As such, this step simply identifies the pivot positions. In each of the given matrices, the leftmost column has a nonzero entry, so it is a pivot column. The topmost (row one) position in this column is a pivot position. Returning to step 1 with the submatrix containg "all rows beneath the pivot" (as required in step 4) in this case just means the second row:

$$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & \# \end{bmatrix}, \quad \text{and} \begin{bmatrix} 0 & \# & \star \end{bmatrix}.$$

The leftmost nonzero entry in the second row is therefore a pivot position. The pivot positions of matrices (2.3.1) are boxed below.

$$\begin{bmatrix} \boxed{\#} & \star & \star \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \boxed{\#} & \star & \star \\ 0 & 0 & \boxed{\#} \end{bmatrix}, \quad \text{and} \begin{bmatrix} \boxed{\#} & \star & \star \\ 0 & \boxed{\#} & \star \end{bmatrix}$$

Step 2 of the row reduction algorithm requires that each pivot is nonzero. This is already the case so no action is required.

Step 3 of the row reduction algorithm requires all entries below a pivot (and in the same column as the pivot) be zero. In all three matrices, the only pivot with entries directly below it are those in the 1,1-entry, and there are zeros beneath them in each case, so no action is required.

Step 4 of the row reduction algorithm sends the algorithm back to the first step. It does not, by itself, provide for any changes to the matrix.

What we can glean from analysis of the algorithm is that in row echelon form,

1. every pivot is to the right of the pivots in rows above it, and
2. all rows of zeros are below all rows with nonzero entries.

Both of these facts are immediate consequences of the algorithm. When a pivot is identified, all entries below it in that column are made zero, so when returning to step 1 with the submatrix containing all rows beneath the pivot, the column with the previously identified pivot contains only zeros. The leftmost nonzero column remaining must be to the right! Requiring that the pivot be nonzero ensures that no row of zeros can appear above a row with nonzeros. As long as these two requirements are satisfied, steps 2 and 3 (the only ones of the first 4 that cause any change to a matrix) are satisfied, so the matrix is in row echelon form.

Noticing that a pivot will always be the leftmost nonzero entry in a row makes determining whether a matrix is in row echelon form a simple matter. Identify the leftmost nonzero entry of each row. These are the pivots. Make sure they are all to the right of the ones above them. Then check to make sure any rows of zeros are at the bottom.

Given this description of row echelon form matrices, there are four row echelon forms for 2×3 matrices besides those in (2.3.1). Can you find them? Answer on page 69. Try any other arrangement of 0s, $\star$s, and #s such as

$$\begin{bmatrix} \# & \# & \star \\ 0 & \# & 0 \end{bmatrix} \text{ or } \begin{bmatrix} \star & \star & \star \\ 0 & \# & \star \end{bmatrix}$$

and you will see that it is either a special case of one these seven forms or there is a substitution of numbers for which the matrix is not in row echelon form. Can you show this is true for these two matrices? Answer on page 69.

A linear system with at least one solution is called **consistent**, perhaps deriving from the fact that the equations do not contradict one another. A linear system with no solution is called **inconsistent**. If a row echelon form of its augmented matrix has a pivot in the last column, the linear system is inconsistent. Indeed, a row containing a pivot in the last column has the form $\begin{bmatrix} 0 & 0 & \cdots & 0 & \# \end{bmatrix}$, which translates to the equation $0 = \#$ (zero equals a nonzero number), which of course is nonsense. Otherwise it is consistent.

For a consistent system (no row echelon form matrix has a pivot in the rightmost column), the corresponding system can be solved by back-substitution. The pivots can also be used to determine the number of solutions. If a row echelon form of the system's augmented matrix has a pivot in each column (beside the rightmost), the linear system has exactly one solution. Otherwise it has infinitely many solutions. The second case is worth closer inspection.

All of the following augmented matrices are in row echelon form and are pivot-free in the rightmost column plus at least one other column.

$$\begin{bmatrix} 1 & 3 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & 5 & -2 & -1 \\ 0 & 5 & -4 & 0 \end{bmatrix} \quad \begin{bmatrix} -4 & 1 & 5 & -7 & 3 \\ 0 & 0 & 5 & 1 & -1 \end{bmatrix}$$

This is enough to know they all have infinitely many solutions, but writing down those solutions still takes a little work.

The linear system represented by the first matrix is

$$x_1 + 3x_2 = 5,$$

a single equation in two variables. Solving for x_1, this system has solutions of the form $x_1 = 5 - 3x_2$ and x_2 arbitrary. This implies that we are free to choose any value for x_2 and use the relation $x_1 = 5 - 3x_2$ to determine x_1. For example, we may let $x_2 = 7$, forcing $x_1 = 5 - 3(7) = -16$. So $-16, 7$ is a solution. In this way, the equation $x_1 = 5 - 3x_2$ (or equivalently the equation $x_1 + 3x_2 = 5$) identifies all the solutions of the system, so we could write $\{x_1 = 5 - 3x_2\}$ as the solution set.

Looking back, it would have been just as well to solve for x_2, yielding $x_2 = \frac{5-x_1}{3}$. This formulation would suggest we are free to choose the value of x_1 and use the formula to determine x_2. Doing so makes the solution set $\left\{x_2 = \frac{5-x_1}{3}\right\}$.

Either variable may be treated as arbitrary, giving what appear to be two different solutions. If this makes you feel a little uneasy, you are not alone. Even if we were never to have seen the second solution, the first one may seem a little unsatisfying. We have a formula for one variable and an implicit understanding that the other is free to take on any value. Perhaps a more satisfying way to write down the set of all solutions is to return to the variable x_2, for which we are free to assign any arbitrary value and let it be r (for **r**-bitrary). Doing this, we have $x_2 = r$ and $x_1 = 5 - 3r$ for a solution set. In the spirit of linear algebra, this solution can be expressed in matrix notation as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 - 3r \\ r \end{bmatrix}$$

or better still as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix} + r \begin{bmatrix} -3 \\ 1 \end{bmatrix}.$$

Take a moment to verify that these matrix representations are equivalent to $x_2 = r$ and $x_1 = 5 - 3r$. The last presentation of the solution should feel most satisfying. It does not favor one variable over the other and gives an explicit formula for the value of each variable. This form of the solution is called **parametric vector form**. Returning to the variable x_1 and setting it equal to r leads to a similar parametric vector form of the solution. Can you find it? Answer on page 69.

Crumpet 15: Particular and Homogeneous Solutions

The solution of the linear system

$$x_1 + 3x_2 = 5$$

can be written as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_p + \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_h$$

where $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_p$, called the particular solution, is any one (particular) solution of the system and $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_h$, called the homogeneous solution, is the solution of the corresponding homogeneous system, $x_1 + 3x_2 = 0$. For example, $\begin{bmatrix} 5 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ \frac{5}{3} \end{bmatrix}$, and $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ are valid particular solutions. The homogeneous system has solution $x_1 = -3x_2$, so any solution that includes all instances where the first variable is -3 times the second, as in $r\begin{bmatrix} -3 \\ 1 \end{bmatrix}$, $r\begin{bmatrix} 1 \\ -\frac{1}{3} \end{bmatrix}$, or $r\begin{bmatrix} -3\sqrt{2} \\ \sqrt{2} \end{bmatrix}$, is a valid homogeneous solution. The set of all sums of one particular solution and one homogeneous solution is the **general solution** of the linear system. See section 5.1.

Putting the second matrix,

$$\begin{bmatrix} -3 & 5 & -2 & -1 \\ 0 & 5 & -4 & 0 \end{bmatrix},$$

into reduced row echelon form will facilitate writing the solution of the corresponding linear system. Subtracting row 2 from row 1 and scaling each row appropriately yields reduced row echelon form

$$\begin{bmatrix} 1 & 0 & -2/3 & 1/3 \\ 0 & 1 & -4/5 & 0 \end{bmatrix}$$

(you should verify this) suggesting that the simplest way to write the solution is

$$x_1 = \frac{1}{3} + \frac{2}{3}x_3; \quad x_2 = \frac{4}{5}x_3.$$

This solution further suggests that we let x_3 be **r**-bitrary and write the solution as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 0 \\ 0 \end{bmatrix} + r \begin{bmatrix} 2/3 \\ 4/5 \\ 1 \end{bmatrix}.$$

For a solution without fractions, we can let $r = 10 + 15s$ (after all, r is arbitrary!), which gives

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &= \begin{bmatrix} 1/3 \\ 0 \\ 0 \end{bmatrix} + (10 + 15s) \begin{bmatrix} 2/3 \\ 4/5 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1/3 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 20/3 + 10s \\ 8 + 12s \\ 10 + 15s \end{bmatrix} \\ &= \begin{bmatrix} 7 \\ 8 \\ 10 \end{bmatrix} + s \begin{bmatrix} 10 \\ 12 \\ 15 \end{bmatrix} \end{aligned}$$

where s is the arbitrary variable. There will always be various ways to write the parametric form of the solution of a system with infinitely many solutions.

Putting the third matrix into reduced row echelon form will facilitate writing down its solution as well.

$$\begin{bmatrix} -4 & 1 & 5 & -7 & 3 \\ 0 & 0 & 5 & 1 & -1 \end{bmatrix}$$

is reduced by subtracting the second row from the first and then scaling the rows appropriately. The resulting reduced row echelon form is

$$\begin{bmatrix} 1 & -\frac{1}{4} & 0 & 2 & -1 \\ 0 & 0 & 1 & \frac{1}{5} & -\frac{1}{5} \end{bmatrix}$$

from which we conclude that $x_1 = -1 + \frac{1}{4}x_2 - 2x_4$ and $x_3 = -\frac{1}{5} - \frac{1}{5}x_4$. Variables x_1 and x_3 are easily written in terms of x_2 and x_4, suggesting we allow x_2 and x_4 to be arbitrary. Letting $x_2 = r$ and $x_4 = s$, it follows that $x_1 = -1 + \frac{1}{4}r - 2s$ and $x_3 = -\frac{1}{5} - \frac{1}{5}s$, so

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ -\frac{1}{5} \\ 0 \end{bmatrix} + r \begin{bmatrix} \frac{1}{4} \\ 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 0 \\ 0 \\ -\frac{1}{5} \\ 1 \end{bmatrix}.$$

Two variables may be set arbitrarily this time, a consequence of the fact that we have four variables and only two pivots. Each variable column without a pivot gives a variable that may be set arbitrarily, what are known as **free variables**. Variables represented by columns with pivots are called **basic variables**. Other forms of the solution may be obtained by letting different pairs of variables be the arbitrary ones or by making substitutions for the arbitrary r and s in the above solution, such as $r = 4t$ and $s = 4 + 5u$. Can you write the solution that results from this substitution? Answer on page 69.

The observations that free variables lead to infinitely many solutions and a pivot in the rightmost column of an augmented matrix leads to no solution justify the existence and uniqueness theorem for linear systems.

Theorem 1. [Existence and Uniqueness] *A linear system is consistent if and only if the rightmost column of its augmented matrix representation is not a pivot column. Furthermore, a consistent system will have (a) exactly one solution if it admits no free variables; and (b) infinitely many solutions if it admits at least one free variable.*

Thus the nature of the solution set for any linear system can be determined from a row echelon form of its associated augmented matrix. In problems where this is the entire question at hand, the row echelon form suffices, and can save a bit of time compared to using the reduced row echelon form. The reduced row echelon form, being a row echelon form itself, can be used for this purpose too, but better serves as a place from which to write down the solutions of the system. Thus in problems where the solution set for a linear system is needed, it is usually worth the time and effort to find the reduced row echelon form.

Reduced row echelon form requires that the entries above pivots are zero (step 5) and that each pivot is 1 (step 6), so reduced row echelon form matrices are row echelon form matrices with these extra two properties. The reduced row echelon form of a 2×3 matrix will take one of the following seven forms, for example.

$$\begin{bmatrix} 1 & \star & \star \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & \star \\ 0 & 1 & \star \end{bmatrix}, \begin{bmatrix} 1 & \star & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & \star \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

It is not by mistake that we refer to ***a*** row echelon form or ***the*** reduced row echelon form of a matrix. Row echelon form is not unique, but reduced row echelon form is unique (see crumpet 23 on page 160).

SageMath

If `M` is a matrix in SageMath, then `M.echelon_form()` returns a row echelon form and `M.rref()` returns the reduced row echelon form. The following code computes a row echelon form and the reduced echelon form of

$$M = \begin{bmatrix} -480 & -340 & -110 & -110 & 100 \\ 242 & 146 & 54 & 54 & -60 \\ -721 & -673 & -277 & -152 & 155 \\ 968 & 809 & 316 & 191 & -215 \\ -1039 & -882 & -268 & -268 & 170 \end{bmatrix}.$$

```
M = matrix(5,5,[-480,-340,-110,-110,100,242,146,54,54,-60,-721,-673,-277,
                -152,155,968,809,316,191,-215,-1039,-882,-268,-268,170])
print("M =");print(M); print()
print("Row echelon form:");print(M.echelon_form()); print()
print("Reduced row echelon form:");print(M.rref())
```

40 The output of the code is

```
M =
[ -480  -340  -110  -110   100]
[  242   146    54    54   -60]
[ -721  -673  -277  -152   155]
[  968   809   316   191  -215]
[-1039  -882  -268  -268   170]

Row echelon form:
[  1  13  87 462  45]
[  0  25  25 150  25]
[  0   0 125 625  50]
[  0   0   0 750 150]
[  0   0   0   0   0]

Reduced row echelon form:
[   1    0    0    0 -2/5]
[   0    1    0    0  2/5]
[   0    0    1    0 -3/5]
[   0    0    0    1  1/5]
[   0    0    0    0    0]
```

Key Concepts

basic variable a variable corresponding to the column of an augmented matrix with a pivot.

free variable a variable corresponding to the column of an augmented matrix with no pivot.

consistent (linear system) having at least one solution.

inconsistent (linear system) having no solution.

existence and uniqueness of solutions (for linear systems) see theorem 1.

parametric vector form a linear combination of vectors.

reduced row echelon form of a matrix is unique.

Exercises

1. The augmented matrix for a linear system is given. (i) Determine whether it is in row echelon form. If it is, state (ii) whether the associated linear system is consistent or inconsistent, and (iii) how many solutions it has.

(a) $\begin{bmatrix} 1 & 4 & 0 & 0 & 3 & -1 \\ 0 & 0 & 3 & 0 & 5 & -6 \\ 0 & 0 & 0 & -1 & 2 & 3 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 & -8 \\ 0 & 1 & 5 \end{bmatrix}$ [𝔸]-349

(c) $\begin{bmatrix} 2 & 4 & 4 & 1 & 1 & 3 \\ 0 & 0 & -4 & 4 & -2 & -4 \\ 0 & -5 & 0 & 3 & 5 & 3 \\ 0 & 0 & 0 & 0 & -5 & 4 \end{bmatrix}$ [𝔸]-349

(d) $\begin{bmatrix} 3 & 0 & 2 & -8 \\ 0 & 1 & 1 & 9 \\ 0 & 0 & 0 & 6 \end{bmatrix}$ [𝔸]-349

(e) $\begin{bmatrix} 6 & -5 & 4 \\ 0 & 0 & -3 \end{bmatrix}$

(f) $\begin{bmatrix} 1 & 0 & 0 & -5 \\ 5 & -3 & 0 & -1 \\ -3 & 4 & -1 & -5 \end{bmatrix}$

(g) $\begin{bmatrix} 3 & -2 & 5 & 6 & 7 \\ 0 & -2 & 0 & 7 & -4 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 18 & 37 \end{bmatrix}$

(h) $\begin{bmatrix} 5 & -4 & 8 & 2 \\ 0 & -3 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ [𝔸]-349

2. The augmented matrix for a linear system is given. (i) Determine whether it is in reduced row echelon form. If it is, (ii) find the solution set, in parametric vector form if infinite.

(a) $\begin{bmatrix} 1 & 7 & 4 \\ 0 & 1 & 0 \end{bmatrix}$

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

(c) $\begin{bmatrix} 1 & 6 & 0 & -2 & 0 & 5 \\ 0 & 0 & 1 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 0 & 6 \\ 0 & 0 & 1 & -2 \\ 0 & 1 & 0 & 4 \end{bmatrix}$ [𝔸]-349

(e) $\begin{bmatrix} 0 & 1 & 0 & 6 & 0 \\ 0 & 0 & 1 & -8 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

(f) $\begin{bmatrix} 1 & 0 & -9 & 0 & 4 \\ 0 & 1 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}$ [𝔸]-349

3. The augmented matrix for a linear system is given. (i) Is the system homogeneous? If yes, (ii) find the solution set, in parametric vector form if infinite.

(a) $\begin{bmatrix} 1 & 0 & 6 & 0 \\ 0 & 1 & -8 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 4 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & 0 & -2 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ [𝔸]-349

(d) $\begin{bmatrix} 3 & 0 & 4 & 2 \\ 0 & 2 & -5 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ [𝔸]-349

4. The coefficient matrix for a homogeneous linear system is given. Find the solution set. If the solution set is infinite give your answer in parametric vector form.

(a) $\begin{bmatrix} 1 & 6 & -2 \\ 2 & 10 & -3 \\ 3 & 16 & -5 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & -7 \\ 0 & -2 & 7 \end{bmatrix}$ [𝕊]-291

(c) $\begin{bmatrix} 6 & 7 & -3 \\ 0 & 4 & 2 \\ 0 & 0 & -1 \end{bmatrix}$ [𝔸]-349

(d) $\begin{bmatrix} 3 & -6 & 8 \\ -6 & 12 & -16 \\ 1 & -2 & \frac{8}{3} \end{bmatrix}$

(e) $\begin{bmatrix} 1 & 4 & 8 \\ -3 & -10 & -20 \\ 2 & 8 & 15 \end{bmatrix}$

(f) $\begin{bmatrix} 2 & 1 & -6 \\ -4 & -2 & 12 \\ 1 & \frac{1}{2} & -3 \end{bmatrix}$ [𝔸]-349

5. Use row reduction and a reduced row echelon form matrix to find all the eigenvectors. All eigenvalues for the matrix are given.

   (a) $\begin{bmatrix} 12 & -8 \\ 15 & -10 \end{bmatrix}; \lambda = 0, 2$

   (b) $\begin{bmatrix} 40 & 66 \\ -22 & -37 \end{bmatrix}; \lambda = 7, -4$ [𝔸]-349

   (c) $\begin{bmatrix} 6 & -4 & 16 \\ 3 & -7 & 4 \\ -6 & 2 & -14 \end{bmatrix}; \lambda = -6, -3$ [𝕊]-291

   (d) $\begin{bmatrix} -7 & -2 & -1 \\ 20 & 6 & 4 \\ -5 & -2 & -3 \end{bmatrix}; \lambda = -2, 0$

6. Using $0, \#, \star$ notation, list all the posible row echelon forms for a 3×3 matrix.

7. What can you say about the existence and uniqueness of solutions of the system associated with the described matrix?

   (a) 4×7 coefficient matrix with 4 pivots

   (b) 4×7 augmented matrix with 4 pivots

   (c) 7×4 coefficient matrix with 4 pivots

   (d) 7×4 augmented matrix with 4 pivots

8. What conditions on the pivots of an augmented matrix would ensure the associated system had a unique solution?

9. What conditions on the pivots of a coefficient matrix would ensure the associated system had a unique solution?

10. Alex and Bianca are solving the same system of equations. For the reduced row echelon form of the augmented matrix, Alex got

$$\begin{bmatrix} 1 & 0 & \star & \star \\ 0 & 1 & \star & \star \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and Bianca got

$$\begin{bmatrix} 1 & \star & 0 & \star \\ 0 & 0 & 1 & \star \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

[𝔸]-349

   (a) Is the system consistent according to Alex's work? According to Bianca's?

   (b) How many solutions does the system have according to Alex's work? According to Bianca's?

   (c) How many free variables does the system have according to Alex's work? According to Bianca's?

   (d) Can they both be right assuming the columns of their augmented matrices both hold the coefficients of x, y, z in that order?

   (e) Suppose the columns of Alex's augmented matrix hold the coefficients of x, y, z in that order, but Bianca's columns hold the coefficients in the order x, z, y? Now can they both be right?

11. The reduced row echelon forms for the coefficient matrices of two linear systems of three equations in three variables are

$$\begin{bmatrix} 1 & \star & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & \star & \star \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

so both systems have infinitely many solutions. Is it possible the systems are the same? Explain.

12. A system of linear equations with fewer equations than unknowns is sometimes called underdetermined. Suppose such a system were consistent. Explain why the system must have an infinite number of solutions. [𝔸]-349

13. A system of linear equations with more equations than unknowns is sometimes called overdetermined. Can such a system be consistent? Illustrate your answer with an example (if yes) or an argument as to why not (if no). [𝔸]-349

14. 41 Use SageMath to find (i) a row echelon form (but not reduced row echelon form) and (ii) the reduced row echelon form of

$$\begin{bmatrix} 2049 & -4548 & -511 & -5177 & 6023 \\ -4526 & 10252 & 916 & 11438 & -13292 \\ -6947 & 15538 & 1740 & 17601 & -20614 \\ -1388 & 2866 & 263 & 3166 & -3697 \\ -5781 & 12812 & 1211 & 14321 & -16671 \end{bmatrix}$$

[𝕊]-292

15. 42 Use SageMath to find the reduced row echelon form of the augmented matrix, and write down the solution of the corresponding linear system. Assume the columns represent the variables x_1, x_2, x_3, x_4, x_5 in that order.

$$\begin{bmatrix} 27 & -36 & -4 & 2 & 4 & 58 \\ 12 & -16 & -2 & 1 & 2 & 27 \\ 15 & -20 & -4 & 2 & 4 & 42 \\ -3 & 4 & -2 & 1 & 2 & 7 \end{bmatrix}$$

16. 43 Suppose the matrix in question 15 is the coefficient matrix of a homogeneous linear system with variables $x_1, x_2, x_3, x_4, x_5, x_6$ and repeat the exercise.

Answers

which is which? The three matrices

$$\begin{bmatrix} 7 & 10 & -3 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 7 & 0 & 0 \\ 0 & 10 & -3 \end{bmatrix}, \quad \text{and} \begin{bmatrix} 7 & 10 & 0 \\ 0 & 0 & -3 \end{bmatrix}$$

have associated linear systems with infinitely many solutions, one solution, and no solution, respectively.

other row echelon forms The other four possible row echelon forms for 2×3 matrices are

$$\begin{bmatrix} 0 & \# & \star \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & \# & \star \\ 0 & 0 & \# \end{bmatrix}, \begin{bmatrix} 0 & 0 & \# \\ 0 & 0 & 0 \end{bmatrix}, \text{ and } \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

cases where the first variable does not appear in either of the equations. The last two forms, where neither variable appears in either equation, have arguable applicability.

substitution The matrix $\begin{bmatrix} \# & \# & \star \\ 0 & \# & 0 \end{bmatrix}$ is in row echelon form for any substitution of numbers, but is a special case of $\begin{bmatrix} \# & \star & \star \\ 0 & \# & \star \end{bmatrix}$. The matrix $\begin{bmatrix} \star & \star & \star \\ 0 & \# & \star \end{bmatrix}$ is not in row echelon form for the substitution $\begin{bmatrix} 0 & 0 & \star \\ 0 & \# & \star \end{bmatrix}$. The pivot in the second column cannot be zero.

parametric vector form Setting $x_1 = r$ yields $x_2 = \frac{5-r}{3}$, which is equivalent to

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} r \\ \frac{5-r}{3} \end{bmatrix}$$

or

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{5}{3} \end{bmatrix} + r \begin{bmatrix} 1 \\ -\frac{1}{3} \end{bmatrix}$$

To make it look a little more like the previous solution, let $r = -3s$, which gives

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{5}{3} \end{bmatrix} + s \begin{bmatrix} -3 \\ 1 \end{bmatrix}.$$

substitution 2 The solution with substitution is

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} &= \begin{bmatrix} -1 \\ 0 \\ -\frac{1}{5} \\ 0 \end{bmatrix} + 4t \begin{bmatrix} \frac{1}{4} \\ 1 \\ 0 \\ 0 \end{bmatrix} + (4 + 5u) \begin{bmatrix} 0 \\ 0 \\ -\frac{1}{5} \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ 0 \\ -\frac{1}{5} \\ 0 \end{bmatrix} + \begin{bmatrix} t \\ 4t \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -\frac{4}{5} - u \\ 4 + 5u \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ 0 \\ -1 \\ 4 \end{bmatrix} + t \begin{bmatrix} 1 \\ 4 \\ 0 \\ 0 \end{bmatrix} + u \begin{bmatrix} 0 \\ 0 \\ -1 \\ 5 \end{bmatrix} \end{aligned}$$

Chapter 3

Matrix Algebra

3.1 Properties of Matrix Operations

Background

If you ever thought that algebra should be renamed "find x" or "how to solve equations", you are not alone. The study of algebra is largely concerned with solving equations. Linear algebra is considerably less concerned with solving equations, but it is still an important feature of the subject. There are many similarities between the rules that govern the manipulation of algebraic expressions involving real numbers and those governing the manipulation of expressions involving matrices, but there are significant differences, all worthy of recording. First, a few words about the arithmetic of real numbers.

During the late 19^{th} century, the European mathematics community set to the task of answering foundational questions about arithmetic. They debated the questions of how to define the natural numbers and the real numbers; how to define the operations of addition and multiplication; and just as importantly to what extent such definitions were useful. The German mathematician Hermann Günther Grassmann (Graßmann) is generally credited with sparking the debate by showing that properties of the natural numbers that to that point had simply been taken for granted (such as the fact that $a + b = b + a$) could be proved from simpler principles. After a number of developments, the Italian mathematician Giuseppe Peano published his *Arithmetices Principia* [20] (Principles of Arithmetic, 1889) summarizing work to that point and adding his stamp in the form of his five axioms defining the natural numbers.

Crumpet 16: Foundations of Analysis

In 1951 Edmund Landau published the first edition of his *Grundlagen der Analysis* (Foundations of Analysis [15], available at https://b-ok.cc/book/2863641/855790, accessed Feb 9, 2021) where, based on the work of Giuseppe Peano, Richard Dedekind, Augustin-Louis Cauchy and others, he develops the arithmetic of whole, rational, irrational and complex numbers in a single volume. Peano's five axioms defining the natural numbers appear on page 2 as follows.

Axiom 1: 1 *is a natural number.*

That is, our set is not empty; it contains an object called 1 (read "one").

Axiom 2: *For each x there exists exactly one natural number, called the successor of x, which will be denoted x'.*

In the case of complicated natural numbers x, we will enclose in parentheses the number whose successor is to be written down, since otherwise ambiguities might arise. We will do the same, throughout this book, in the case of $x + y$, xy, $x - y$, $-x$, x^y,etc.

Thus, if

$$x = y$$

then

$$x' = y'.$$

Axiom 3: *We always have*

$$x' \neq 1.$$

That is, there exists no number whose successor is 1.

Axiom 4: *If*

$$x' = y'$$

then

$$x = y.$$

That is, for any given number there exists either no number or exactly one number whose successor is the given number.

Axiom 5 (Axiom of Induction): *Let there be given a set $\mathfrak{R}$ of natural numbers, with the following properties:*

I) 1 *belongs to* $\mathfrak{R}$.

II) *If* x *belongs to* $\mathfrak{R}$ *then so does* x'.

Then $\mathfrak{R}$ *contains all the natural numbers.*

Kurt Friedrich Gödel's incompleteness theorems, published in 1931 [9], essentially concluded the debate. He proved that any consistent axiomatic system sufficient to describe arithmetic on the natural numbers (including Peano's five axioms) will admit statements that cannot be proven nor disproven from within the system. Despite this deficiency, Peano's axioms are sufficient to define natural numbers and prove the familiar properties of the operations of addition and multiplication. With the comfort of knowing these facts rest on solid foundation, we will assume their veracity. That is not to say we will simply take them for granted, however.

$1+3$ equals 4, and $3+1$ equals 4. $7+9$ equals 16, and $9+7$ equals 16. The more general statement that $a+b = b+a$ for any numbers a and b is called the commutative property for addition of real numbers. Though this property is one of the basic principles that can be proven based on even more basic principles, we will take the viewpoint that it had to turn out this way! The counting numbers, $1, 2, 3, \ldots$, represent how many of a thing we have (quantity). This is a fact engrained in our minds as we learn to count—at a very young age. Addition models what happens when two quantities are merged—another concept engrained in our minds early on in our mental development. If you add three apples to a basket initially holding one apple, afterward it will contain four apples. This merger is modeled by the statement $1 + 3 = 4$ (one apple plus three more apples gives you four apples). Similarly if you add one apple to a basket initially holding three apples, afterward it will contain four apples. This merger is modeled by the statement $3 + 1 = 4$. The fact that a pair of natural numbers can be added in either order with the same result is simply the way natural numbers work. Any mathematical axiom, theorem, or proof suggesting otherwise is simply not the number system we were taught as youngsters.

Crumpet 17: An Interesting Addition Table

If we completely abandon the usual notions of natural numbers (that there are infinitely many of them and that they represent quantities, for example), addition might be defined as follows.

+	1	2	3	4	5	6
1	2	6	4	5	3	1
2	6	1	5	3	4	2
3	5	4	6	2	1	3
4	3	5	1	6	2	4
5	4	3	2	1	6	5
6	1	2	3	4	5	6

This sytem of addition retains some of the familiar notions of arithmetic such as associativity and the existence of an identity (can you tell which symbol acts as the identity?), but not commutativity. According to this table, $1 + 3 = 4$

but $3+1=5$! Addition in this system is not commutative, but that does not make this number system "wrong". It simply means this system, an example of a finite group, does not represent the numbers as we commonly understand them. It makes a poor model for the measurement of quantity.

Some Properties of Matrices

Table 3.1: Some Properties of Real Numbers

For all real numbers a, b, c	
a. $a+b=b+a$	Commutative property for addition
b. $(a+b)+c=a+(b+c)$	Associative property for addition
c. $a+0=a=0+a$	Additive identity
d. $a(bc)=(ab)c$	Associative property for multiplication
e. $a(b+c)=ab+ac$	Distributive property
f. $1 \cdot a=a=a \cdot 1$	Multiplicative identity

Table 3.1 summarizes the properties of real numbers of interest to our study of linear algebra, each of which has a matrix analog as shown in the following theorem. An $m \times n$ matrix whose entries are all zero is called a **zero matrix** and is denoted $0_{m \times n}$ or just 0 when its size is discernible through context.

Theorem 2. [Algebraic Properties of Matrices, Part 1] *For all matrices A, B, C*

1. $A+B=B+A$ *(commutative property for addition)*
2. $(A+B)+C=A+(B+C)$ *(associative property for addition)*
3. $A+0=A=0+A$ *(additive identity)*
4. $A(BC)=(AB)C$ *(associative property for multiplication)*
5. $A(B+C)=AB+AC$ *(left distributive property)*
6. $IA=A=AI$ *(multiplicative identity)*

whenever the indicated operations are defined.

Claim 1 can be proven by noting that

$$(A+B)_{i,j}=A_{i,j}+B_{i,j}$$

and

$$(B+A)_{i,j}=B_{i,j}+A_{i,j}$$

by definition of matrix addition. Since $A_{i,j}$ and $B_{i,j}$ are numbers, the commutative property for addition of real numbers allows us to use the fact that $A_{i,j}+B_{i,j}=B_{i,j}+A_{i,j}$ to deduce that $(A+B)_{i,j}=(B+A)_{i,j}$. Since the entries of $A+B$ and $B+A$ are equal, the matrices are equal.

In more words than symbols, we might argue as follows. The i,j-entries of $A+B$ and $B+A$ are calculated by adding the same two entries of A and B only in different orders. Since addition of real numbers (entries) is commutative, the i,j-entries are equal. Hence $A+B=B+A$.

In more symbols than words, a third way to see that theorem 2 claim 1 is true, consider the following computation.

$$
\begin{aligned}
A+B &= \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{bmatrix} + \begin{bmatrix} B_{1,1} & B_{1,2} & \cdots & B_{1,n} \\ B_{2,1} & B_{2,2} & \cdots & B_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m,1} & B_{m,2} & \cdots & B_{m,n} \end{bmatrix} \\
&= \begin{bmatrix} A_{1,1}+B_{1,1} & A_{1,2}+B_{1,2} & \cdots & A_{1,n}+B_{1,n} \\ A_{2,1}+B_{2,1} & A_{2,2}+B_{2,2} & \cdots & A_{2,n}+B_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1}+B_{m,1} & A_{m,2}+B_{m,2} & \cdots & A_{m,n}+B_{m,n} \end{bmatrix} \\
&= \begin{bmatrix} B_{1,1}+A_{1,1} & B_{1,2}+A_{1,2} & \cdots & B_{1,n}+A_{1,n} \\ B_{2,1}+A_{2,1} & B_{2,2}+A_{2,2} & \cdots & B_{2,n}+A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m,1}+A_{m,1} & B_{m,2}+A_{m,2} & \cdots & B_{m,n}+A_{m,n} \end{bmatrix} \\
&= B+A
\end{aligned}
$$

In essence, the fact that matrices obey the rule $A + B = B + A$ follows directly from the commutative property for addition of real numbers and the fact that addition of matrices is computed component-wise. The rest of the claims in theorem 2 can also be justified by use of the corresponding property for real numbers and careful application of the definitions of matrix addition and multiplication.

As noted in section 1.3 matrix multiplication is not commutative. The familiar commutative property for multiplication of real numbers, $ab = ba$, does not have a matrix analog. Thus it is important to point out, for example, that the distributive property for matrices holds for both left-multiplication (as in theorem 2 claim 5) and right-multiplication. Additionally, the interplay between scalar multiplication and both matrix addition and matrix multiplication must be documented, as seen in the following theorem.

Theorem 3. [Algebraic Properties of Matrices, Part 2] *For all matrices A, B, C and scalars r, s*

1. $r(A + B) = rA + rB$ *(distributivity of a scalar over matrices)*
2. $(r + s)A = rA + sA$ *(distributivity of a matrix over scalars)*
3. $(rs)A = r(sA)$ *(associativity of multiplication between two scalars and one matrix)*
4. $r(AB) = (rA)B = A(rB)$ *(associativity of multiplication between a scalar and two matrices)*
5. $(B + C)A = BA + CA$ *(right distributive property)*

whenever the indicated operations are defined.

Again these claims can be justified by use of the corresponding property for real numbers and careful application of the definitions of matrix addition and multiplication.

One final theorem for this section contains a list of identities concerning the matrix transpose and inverse. Claims concerning only the transpose can be proven by comparing i, j-entries as before while those concerning the inverse are most easily proven without reference to individual entries.

Theorem 4. [Algebraic Properties of Matrices, Part 3] *For all matrices A and B and scalar r*

1. $(A^T)^T = A$ *(transpose of the transpose)*
2. $(rA)^T = rA^T$ *(transpose of a scalar multiple)*
3. $(rA)^{-1} = \frac{1}{r}A^{-1}$ *(inverse of a scalar multiple)*
4. $(A + B)^T = A^T + B^T$ *(transpose of a sum)*
5. $(AB)^T = B^T A^T$ *(transpose of a product)*

6. $(A^{-1})^{-1} = A$ *(inverse of the inverse)*

7. $(AB)^{-1} = B^{-1}A^{-1}$ *(inverse of a product)*

8. $(A^T)^{-1} = (A^{-1})^T$ *(inverse of the transpose)*

whenever the indicated operations are defined.

To make the point that the claims involving inverses can be proven without reference to entries, consider theorem 4 claim 6. In words it says the inverse of the inverse of a matrix is the matrix itself. Or in other words, if you find the inverse of a matrix, then find the inverse of that matrix you get the original matrix. More in the words of the definition of inverse (matrices A and B are inverses if and only if $AB = BA = I$), any statement about inverses can be rephrased as a statement about a product. To show that two matrices are inverses, it is often best to show that their products, in both orders, each equal the identity. As for claim 6, it suffices to show $AA^{-1} = A^{-1}A = I$ (the inverse of A^{-1} is the matrix B such that $BA^{-1} = A^{-1}B = I$), but that equality is true due to the definition of inverse—end of proof. The claim is more a matter of perspective than a claim of something new.

While a list of 19 claims over 3 theorems may seem a bit overwhelming to digest, there are only a small few that require great attention. The claims of theorem 2 are replicas of properties of real numbers with which you are hopefully familiar. As such, it is the differences between the algebra of numbers and the algebra of matrices that should gain your focus. Primarily there is no commutative property for multiplication of matrices! This has consequences such as the appearance of claim 5 of theorem 3. The right distributive property is not necessary to enumerate separately for real numbers because it follows from the combination of commutativity for multiplication and distributivity of real numbers. The rest of theorem 3 is documentation of what you would probably expect to be true about scalar multiplication.

In addition to the right distributive property, theorem 4 is worth careful scrutiny. It contains facts about the relatively new concepts of inverses and transposes. In particular, notice that (claim 5) the transpose of a product is the product of the transposes in the opposite order! Similarly, (claim 7) the inverse of a product is the product of the inverses in the opposite order. Justification for this claim is requested in exercise 17.

Applications to eigenpairs

As an example of the utility of the properties of theorems 2 through 4, the claim *if* $\mathbf{v}$ *is an eigenvector of* A *associated with value* λ*, so is* $c\mathbf{v}$ of section 1.7, page 43, can now be justified. Assuming $\mathbf{v}$ is an eigenvector of A, we know that $A\mathbf{v} = \lambda\mathbf{v}$ (by definition). To justify the claim, we need to demonstrate that $A(c\mathbf{v}) = \lambda(c\mathbf{v})$:

$$\begin{aligned}
A(c\mathbf{v}) &= c(A\mathbf{v}) && \text{theorem 3 claim 4}\\
&= c(\lambda\mathbf{v}) && \text{definition of eigenpair}\\
&= (c\lambda)\mathbf{v} && \text{theorem 3 claim 3}\\
&= (\lambda c)\mathbf{v} && \text{commutative property for multiplication of real numbers}\\
&= \lambda(c\mathbf{v}) && \text{theorem 3 claim 3}
\end{aligned}$$

Each algebraic manipulation must be supported by one of the theorems or a property of the real numbers.

As a second example of the utility of these theorems, we are also prepared to prove the claim that if $(\lambda, \mathbf{v})$ is an eigenpair for the matrix M, then $(M - \lambda I)\mathbf{v} = \mathbf{0}$, also from section 1.7. Technically the first line of the justification should itself be justified in some general form before it is used. Can you supply such a proof? Answer on page 77.

$$\begin{aligned}
\mathbf{0} &= M\mathbf{v} - M\mathbf{v} && \text{the difference between a matrix and itself is a zero matrix}\\
&= M\mathbf{v} - \lambda\mathbf{v} && \text{definition of eigenpair (substitution of } \lambda\mathbf{v} \text{ for } M\mathbf{v})\\
&= M\mathbf{v} - \lambda(I\mathbf{v}) && \text{definition of identity matrix (substitution of } I\mathbf{v} \text{ for } \mathbf{v})\\
&= M\mathbf{v} - (\lambda I)\mathbf{v} && \text{theorem 3 claim 4}\\
&= (M - \lambda I)\mathbf{v} && \text{theorem 3 claim 5}
\end{aligned}$$

Key Concepts

algebraic properties of matrices see theorems 2, 3, and 4.

zero matrix an $m \times n$ matrix whose entries are all zero, denoted $0_{m\times n}$ or just 0 when its size is discernible through context.

Exercises

1. Illustrate the property by example. Then explain in your own words why it is true.
   (a) $(A+B)+C = A+(B+C)$ (theorem 2 claim 2).
   (b) $A+0 = A$ (theorem 2 claim 3).
   (c) $r(A+B) = rA+rB$ (theorem 3 claim 1).
   (d) $(r+s)A = rA+sA$ (theorem 3 claim 2).
   (e) $(rs)A = r(sA)$ (theorem 3 claim 3). [𝔸]-349
   (f) $(rA)B = A(rB)$ (theorem 3 claim 4).
   (g) $(A^T)^T = A$ (theorem 4 claim 1).
   (h) $(rA)^T = rA^T$ (theorem 4 claim 2).
   (i) $(A+B)^T = A^T+B^T$ (theorem 4 claim 4). [𝔸]-349
2. Justify the property by showing that the i,j-entry of the lefthand side of the equation equals the i,j-entry of the righthand side.
   (a) $A+0 = A$ (theorem 2 claim 3).
   (b) $A(BC) = (AB)C$ (theorem 2 claim 4). [𝕊]-293
   (c) $A(B+C) = AB+AC$ (theorem 2 claim 5).
   (d) $(r+s)A = rA+sA$ (theorem 3 claim 2).
   (e) $(rA)B = A(rB)$ (theorem 3 claim 4).
   (f) $(B+C)A = BA+CA$ (theorem 3 claim 5).
   (g) $(A^T)^T = A$ (theorem 4 claim 1). [𝕊]-293
   (h) $(A+B)^T = A^T+B^T$ (theorem 4 claim 4).
   (i) $(AB)^T = B^TA^T$ (theorem 4 claim 5).
3. Justify the claim by a string of equalities where each equality is supported by a definition or theorem or claim you justify separately.
   (a) $(AB)^{-1} = B^{-1}A^{-1}$ (theorem 4 claim 7). [𝕊]-293
   (b) $(A^T)^{-1} = (A^{-1})^T$ (theorem 4 claim 8).
4. Let $A = \begin{bmatrix} 1 & 0 \\ -3 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & -2 \\ 1 & 5 \end{bmatrix}$.
   (a) Compute AB^T.
   (b) Without any further computation, find BA^T and explain how you got it.
5. Let $A = \begin{bmatrix} 4 & 7 \\ 3 & 5 \end{bmatrix}$. [𝔸]-350
   (a) Compute A^{-1}.
   (b) Without any further computation, find $(A^T)^{-1}$ and explain how you got it.
6. Let $A = \begin{bmatrix} 11 & 10 & -6 \\ -3 & 6 & 7 \end{bmatrix}$. Calculate $3A+7A$ without calculating $3A$ or $7A$. [𝕊]-293
7. Let $A = \begin{bmatrix} 4 & 7 & -10 \\ -3 & -1 & 7 \end{bmatrix}$. Calculate $2(3A)$ without calculating $3A$.
8. Let $A = \begin{bmatrix} 9 & 6 \\ -2 & 7 \end{bmatrix}$; $B = \begin{bmatrix} -10 & 3 \\ 5 & -8 \end{bmatrix}$; and $C = \begin{bmatrix} -1 & 3 & -11 \\ 1 & -2 & 9 \end{bmatrix}$. Calculate $AC+BC$ without calculating AC or BC.
9. Let $A = \begin{bmatrix} 3 & 4 \\ -4 & -5 \end{bmatrix}$. Calculate $\left(\frac{1}{3}A\right)^{-1}$ without calculating $\frac{1}{3}A$.
10. Let $A = \begin{bmatrix} 9 & 3 & -5 \\ 10 & 11 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} -3 & 5 & -6 \\ -8 & 12 & 0 \end{bmatrix}$. Calculate $(A+B)^T$ without calculating $A+B$.
11. Calculate $(AB)^T$ without calculating AB.
    (a) $A = \begin{bmatrix} 0 & -8 \\ -12 & 5 \end{bmatrix}$; $B = \begin{bmatrix} 10 & -5 \\ 7 & 0 \end{bmatrix}$
    (b) $A = \begin{bmatrix} 9 & 3 \\ -4 & -10 \end{bmatrix}$; $B = \begin{bmatrix} 1 & 9 \\ 2 & -8 \end{bmatrix}$
12. Calculate $(AB)^{-1}$ without calculating AB.
    (a) $A = \begin{bmatrix} -3 & 2 \\ 7 & -5 \end{bmatrix}$; $B = \begin{bmatrix} -8 & -3 \\ 5 & 2 \end{bmatrix}$
    (b) $A = \begin{bmatrix} 4 & 1 \\ 7 & 2 \end{bmatrix}$; $B = \begin{bmatrix} -2 & 3 \\ 1 & -2 \end{bmatrix}$
13. Justify the claim
    $$(A+B)(C+D) = AC+AD+BC+BD$$
    using a series of equalities, each one supported by a theorem.
14. Justify the claim by arguing that the i,j-entry on the lefthand side of the conclusion equals the i,j-entry on the righthand side of the conclusion.
    (a) If $A = B$ then $A-B = 0$. (The conclusion is $A-B = 0$. Use the assumption that $A = B$ in your argument.)
    (b) If $A-B = 0$ then $A = B$. (The conclusion is $A = B$. Use the assumption that $A-B = 0$ in your argument.)
15. Let A be an $m \times n$ matrix. Justify the claim by arguing that the i,j-entry on the lefthand side equals the i,j-entry on the righthand side.
    (a) $A0_{n\times\ell} = 0_{m\times\ell}$ for any positive integer ℓ.
    (b) $0_{\ell\times m}A = 0_{\ell\times n}$ for any positive integer ℓ.
16. Show that, for any matrix A, $-A$ is the additive inverse of A. That is, show
    (a) $-A+A = 0$; and
    (b) $A+(-A) = 0$. [𝔸]-350

    $-A$ has the common meaning $-1 \cdot A$.
17. Justify theorem 4 part 7. [𝔸]-350

Answers

difference of a matrix with itself The general statement that if A is any matrix then $A - A = 0$ can be proven by noting that

$$(A - A)_{i,j} = A_{i,j} - A_{i,j} = 0$$

for all entries $(A - A)_{i,j}$.

3.2 Matrix Equations

The algebraic equation

$$x - 7 = 5$$

is commonly solved by adding 7 to both sides of the equation. The reason this works is because $-7 + 7 = 0$ (7 is the additive inverse of -7) and $x + 0 = x$ (0 is the additive identity). In linear algebra there is an additive identity matrix (theorem 2 claim 3) and there are additive inverses (section 3.1 exercise 16), so analogous equations ought to be solvable similarly. Indeed they are! The matrix equation

$$X - \begin{bmatrix} 7 & 1 \\ -3 & 8 \end{bmatrix} = \begin{bmatrix} 5 & -3 \\ 2 & 4 \end{bmatrix} \tag{3.2.1}$$

can be solved by the same process. Adding $\begin{bmatrix} 7 & 1 \\ -3 & 8 \end{bmatrix}$ to both sides of the equation gives

$$\begin{aligned} X - \begin{bmatrix} 7 & 1 \\ -3 & 8 \end{bmatrix} + \begin{bmatrix} 7 & 1 \\ -3 & 8 \end{bmatrix} &= \begin{bmatrix} 5 & -3 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} 7 & 1 \\ -3 & 8 \end{bmatrix} \\ X + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} &= \begin{bmatrix} 5+7 & -3+1 \\ 2-3 & 4+8 \end{bmatrix} \\ X &= \begin{bmatrix} 12 & -2 \\ -1 & 12 \end{bmatrix} \end{aligned}$$

Substituting $\begin{bmatrix} 12 & -2 \\ -1 & 12 \end{bmatrix}$ for X in equation (3.2.1) yields a true statement, so $\begin{bmatrix} 12 & -2 \\ -1 & 12 \end{bmatrix}$ is a solution.

The slightly more advanced equation

$$8x - 7 = 5$$

is commonly solved by first adding 7 to both sides and then dividing both sides by 8 (or equivalently multiplying both sides by $\frac{1}{8}$, the multiplicative inverse of 8). This might be demonstrated as follows.

$$\begin{aligned} 8x - 7 &= 5 \\ 8x - 7 + 7 &= 5 + 7 \\ 8x &= 12 \\ \frac{8x}{8} &= \frac{12}{8} \\ x &= \frac{3}{2} \end{aligned}$$

This method works because, in addtion to the facts that -7 and 7 are additive inverses and 0 is the additive identity, $\frac{8}{8} = 1$ (8 and $\frac{1}{8}$ are multiplicative inverses) and $1x = x$ (1 is the multiplicative identity). In linear algebra, there are multiplicative identity matrices and there are multiplicative inverses (section 1.6), so analogous equations ought to be solvable similarly. Indeed they are! The matrix equation

$$\begin{bmatrix} 9 & 4 \\ 4 & 2 \end{bmatrix} X - \begin{bmatrix} 7 & 1 \\ -3 & 8 \end{bmatrix} = \begin{bmatrix} 5 & -3 \\ 2 & 4 \end{bmatrix}$$

can be solved by the same process. Adding $\begin{bmatrix} 7 & 1 \\ -3 & 8 \end{bmatrix}$ to both sides of the equation and then (left) multiplying both sides by $\begin{bmatrix} 9 & 4 \\ 4 & 2 \end{bmatrix}^{-1}$ gives

$$\begin{aligned}
\begin{bmatrix} 9 & 4 \\ 4 & 2 \end{bmatrix} X - \begin{bmatrix} 7 & 1 \\ -3 & 8 \end{bmatrix} + \begin{bmatrix} 7 & 1 \\ -3 & 8 \end{bmatrix} &= \begin{bmatrix} 5 & -3 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} 7 & 1 \\ -3 & 8 \end{bmatrix} \\
\begin{bmatrix} 9 & 4 \\ 4 & 2 \end{bmatrix} X &= \begin{bmatrix} 12 & -2 \\ -1 & 12 \end{bmatrix} \\
\begin{bmatrix} 9 & 4 \\ 4 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 9 & 4 \\ 4 & 2 \end{bmatrix} X &= \begin{bmatrix} 9 & 4 \\ 4 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 12 & -2 \\ -1 & 12 \end{bmatrix} \\
\begin{bmatrix} 1 & -2 \\ -2 & \frac{9}{2} \end{bmatrix} \begin{bmatrix} 9 & 4 \\ 4 & 2 \end{bmatrix} X &= \begin{bmatrix} 1 & -2 \\ -2 & \frac{9}{2} \end{bmatrix} \begin{bmatrix} 12 & -2 \\ -1 & 12 \end{bmatrix} \\
\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} X &= \begin{bmatrix} 14 & -26 \\ \frac{-57}{2} & 58 \end{bmatrix} \\
X &= \begin{bmatrix} 14 & -26 \\ \frac{-57}{2} & 58 \end{bmatrix}
\end{aligned}$$

Substituting $\begin{bmatrix} 14 & -26 \\ \frac{-57}{2} & 58 \end{bmatrix}$ for X in the original equation yields a true statement, so $\begin{bmatrix} 14 & -26 \\ \frac{-57}{2} & 58 \end{bmatrix}$ is a solution.

It seems as though the familiar ideas of adding, subtracting, multiplying, and "dividing" both sides of an equation are valid steps in solving matrix equations, but better not to take it for granted on the back of a pair of examples. In adding a matrix to both sides of an equation (or subtracting a matrix from both sides of an equation) in an attempt to solve it, we are using the principle that for matrices A, B, C

$$\text{if } A = B \text{ then } A + C = B + C \tag{3.2.2}$$

whenever the indicated operations are defined. Since the veracity of this proposition is critical to the logical validity of the solutions above, solid proof is warranted.

The principle of equality suggested in exercise 14 of section 3.1 is useful. It says that $A = B$ if and only if $A - B = 0$. That is, if we know that $A = B$, we can safely say that $A - B = 0$. And if we know that $A - B = 0$, we can safely say that $A = B$. Hopefully that sounds logical whether you have completed exercise 14 or not.

Getting back to proposition (3.2.2), note that it begins with the assumption that $A = B$. By the principle of equality we can immediately deduce that $0 = A - B$. Now because $C - C = 0$ for any matrix C and 0 is the additive identity, we can proceed as follows.

$$\begin{aligned}
0 &= A - B \\
&= (A - B) + 0 \\
&= (A - B) + (C - C) \\
&= ((A - B) + C) - C \\
&= (A + (-B + C)) - C \\
&= (A + (C - B)) - C \\
&= ((A + C) - B) - C \\
&= (A + C) + (-B - C) \\
&= (A + C) - (B + C)
\end{aligned}$$

We have used associativity and commutativity for addition of matrices as well as the distributive property. Now that we have $0 = (A + C) - (B + C)$ we conclude that $A + C = B + C$.

Equally critical to the second solution above is the principle that for matrices A, B, C

$$\text{if } A = B \text{ then } CA = CB \tag{3.2.3}$$

whenever the indicated operations are defined. The proof is very similar to the proof of (3.2.2). Starting with $A = B$ allows us to proceed with $0 = A - B$, but this time we employ the fact that $C0 = 0$ for any matrix C (exercise 15 of section 3.1):

$$\begin{aligned} 0 &= A - B \\ &= C(A - B) \\ &= CA - CB \end{aligned}$$

and therefore $CA = CB$. It is also true that

$$\text{if } A = B \text{ then } AC = BC \tag{3.2.4}$$

whenever the indicated operations are defined. Why is this claim needed? Can you justify this claim? Answers on page 84.

Symbolic equations

The need to solve a wholly symbolic equation often arises in the study of mathematics. Principles (3.2.2), (3.2.3), and (3.2.4) are often used to solve such equations. Suppose $XA - I = B$ and we are interested in solving for X, for example. The process is the same as we would use if all the symbols represented numbers. We isolate the X by adding to or multiplying both sides of the equation by appropriate matrices:

$$\begin{aligned} XA - I + I &= B + I \\ XA &= B + I \\ XAA^{-1} &= (B + I)A^{-1} \\ X &= (B + I)A^{-1} \end{aligned}$$

Notice the careful right-multiplication of both sides in the third line. It is not valid to left-multiply one side of the equation while right-multiplying the other. Also, we should note that this solution is only good as long as A is invertible!

The most important equation in linear algebra

For essentialy the entire rest of this textbook, we will be concerned with solving equations of the form

$$M\mathbf{v} = \mathbf{b} \tag{3.2.5}$$

for $\mathbf{v}$. Symbolically the solution is straightforward when M is invertible. Using associativity of multiplication, principle (3.2.3), and the definitions of inverse and identity matrices:

$$\begin{aligned} M^{-1}(M\mathbf{v}) &= M^{-1}\mathbf{b} \\ (M^{-1}M)\mathbf{v} &= M^{-1}\mathbf{b} \\ I\mathbf{v} &= M^{-1}\mathbf{b} \\ \mathbf{v} &= M^{-1}\mathbf{b} \end{aligned} \tag{3.2.6}$$

but understanding it and its ramifications is not. Plus, what if M is not invertible?

On page 51 it was discovered that the product of a matrix A left-multiplied by an elementary matrix E could be calculated one row at a time by noting that each row of EA is the linear combination of the rows of A with coefficients from the corresponding row of E. In symbols, row r of EA can be computed as $(EA)_{r,:} = E_{r,1}A_{1,:} + E_{r,2}A_{2,:} + \cdots + E_{r,n}A_{n,:}$. But this computation holds for any matrix product. Given any matrices B and A where BA is defined—that is, B has the same number of columns as A has rows, say n—row r of BA can be computed as

$$(BA)_{r,:} = B_{r,1}A_{1,:} + B_{r,2}A_{2,:} + \cdots + B_{r,n}A_{n,:}. \tag{3.2.7}$$

This fact is helpful in understanding matrix products such as the one in (3.2.5). For example, suppose the third row of M is twice the first row of M. Then the third row of $\mathbf{b}$ must be twice the first row of $\mathbf{b}$. From (3.2.7)

$$\begin{aligned} \mathbf{b}_{1,:} &= M_{1,1}\mathbf{v}_{1,:} + M_{1,2}\mathbf{v}_{2,:} + \cdots + M_{1,n}\mathbf{v}_{n,:} \\ \mathbf{b}_{3,:} &= M_{3,1}\mathbf{v}_{1,:} + M_{3,2}\mathbf{v}_{2,:} + \cdots + M_{3,n}\mathbf{v}_{n,:} \end{aligned}$$

but $M_{3,:} = 2M_{1,:}$, which means $M_{3,1} = 2M_{1,1},\ M_{3,2} = 2M_{1,2}, \ldots, M_{3,n} = 2M_{1,n}$, so

$$\begin{aligned}\mathbf{b}_{3,:} &= 2M_{1,1}\mathbf{v}_{1,:} + 2M_{1,2}\mathbf{v}_{2,:} + \cdots + 2M_{1,n}\mathbf{v}_{n,:}\\ &= 2(M_{1,1}\mathbf{v}_{1,:} + M_{1,2}\mathbf{v}_{2,:} + \cdots + M_{1,n}\mathbf{v}_{n,:})\\ &= 2\mathbf{b}_{1,:}\end{aligned}$$

As a consequence, for such a matrix M, equation (3.2.5) can only have solutions if the third row of **b** happens to be twice the first row of **b**. There are choices of **b** for which the equation has no solution!

Exercises 9 through 11 of section 1.5 provide evidence that any matrix M in which some row is a linear combination of the others (this is certainly the case for a matrix where the third row is twice the first) will have determinant zero. Formula (1.6.2) suggests that matrices with zero determinant are not invertible. Stringing all this together, it seems the following concepts are interconnected.

- Some row of M is a linear combination of the others.
- M has determinant zero.
- M is not invertible.
- The equation $M\mathbf{v} = \mathbf{b}$ has no solution for certain choices of **b**.

We are not quite ready to prove the connection, but the pieces of the puzzle are falling into place. Taking a slightly different perspective on matrix multiplication will add one more item to the list.

Just as we can imagine the product of two matrices as a collection of linear combinations of the rows of the righthand matrix, we can also imagine the product as a collection of linear combinations of the columns of the lefthand matrix. Thinking generally again, suppose we are given an arbitrary pair of matrices B and A where BA is defined—that is, B has the same number of columns as A has rows, say n. By definition the i,c-entry of BA is $B_{i,1}A_{1,c} + B_{i,2}A_{2,c} + \cdots + B_{i,n}A_{n,c}$. Swapping the order of each product, $(BA)_{i,c} = A_{1,c}B_{i,1} + A_{2,c}B_{i,2} + \cdots + A_{n,c}B_{i,n}$. In particular,

$$\begin{aligned}(BA)_{1,c} &= A_{1,c}B_{1,1} + A_{2,c}B_{1,2} + \cdots + A_{n,c}B_{1,n}\\ (BA)_{2,c} &= A_{1,c}B_{2,1} + A_{2,c}B_{2,2} + \cdots + A_{n,c}B_{2,n}\\ &\vdots\\ (BA)_{m,c} &= A_{1,c}B_{m,1} + A_{2,c}B_{m,2} + \cdots + A_{n,c}B_{m,n}\end{aligned} \tag{3.2.8}$$

Reading these equations together (as columns of numbers) from left to right, they imply that column c of BA (the lefthand sides of the equations) equals $A_{1,c}$ times column 1 of B (the first terms of the righthand sides) plus $A_{2,c}$ times column 2 of B (the second terms of the righthand sides) plus $A_{3,c}$ times column 3 of B (the third terms of the righthand sides), and so on. In symbols,

$$\begin{bmatrix}(BA)_{1,c}\\(BA)_{2,c}\\\vdots\\(BA)_{m,c}\end{bmatrix} = A_{1,c}\begin{bmatrix}B_{1,1}\\B_{2,1}\\\vdots\\B_{m,1}\end{bmatrix} + A_{2,c}\begin{bmatrix}B_{1,2}\\B_{2,2}\\\vdots\\B_{m,2}\end{bmatrix} + \cdots + A_{n,c}\begin{bmatrix}B_{1,n}\\B_{2,n}\\\vdots\\B_{m,n}\end{bmatrix}$$

In other words, column c of BA is a linear combination of the columns of B where the coefficients for the linear combination come from column c of A. In short,

$$(BA)_{:,c} = A_{1,c}B_{:,1} + A_{2,c}B_{:,2} + \cdots + A_{n,c}B_{:,n}. \tag{3.2.9}$$

In the special case of equation (3.2.5),

$$(M\mathbf{v})_{:,c} = \mathbf{v}_{1,c}M_{:,1} + \mathbf{v}_{2,c}M_{:,2} + \cdots + \mathbf{v}_{n,c}M_{:,n}$$

but **v** is a vector (column matrix) so it has only one column. Accordingly, $M\mathbf{v}$ has only one column and we can write

$$M\mathbf{v} = \mathbf{v}_{1,1}M_{:,1} + \mathbf{v}_{2,1}M_{:,2} + \cdots + \mathbf{v}_{n,1}M_{:,n}. \tag{3.2.10}$$

Revisiting the case where the third row of M is twice the first, this means

$$\begin{aligned}\mathbf{b} = M\mathbf{v} &= \mathbf{v}_{1,1}M_{:,1} + \mathbf{v}_{2,1}M_{:,2} + \mathbf{v}_{3,1}(2M_{:,1}) + \cdots + \mathbf{v}_{n,1}M_{:,n} \\ &= (\mathbf{v}_{1,1} + 2\mathbf{v}_{3,1})M_{:,1} + \mathbf{v}_{2,1}M_{:,2} + \cdots + \mathbf{v}_{n,1}M_{:,n}\end{aligned}$$

Letting $\mathbf{v}_{1,1} = -2$, $\mathbf{v}_{3,1} = 1$ and $\mathbf{v}_{j,1} = 0$ for all j not equal to 1 or 3, it turns out $\mathbf{b} = \mathbf{0}$. In other words, the associated homogeneous equation $M\mathbf{v} = \mathbf{0}$ has a solution where $\mathbf{v} \neq \mathbf{0}$, a **nontrivial solution**. Finally, we add

- $M\mathbf{v} = \mathbf{0}$ has a nontrivial solution.

to the list of related concepts. But wait, there's more!

Equation (3.2.8) has the form of a system of linear equations with variables $A_{1,c}, A_{2,c}, \ldots, A_{n,c}$. Can you show that the equation $M\mathbf{v} = \mathbf{b}$ is equivalent to a linear system of equations whose augmented matrix is $\begin{bmatrix} M & \mathbf{b} \end{bmatrix}$ and variables are $\mathbf{v}_{1,1}, \mathbf{v}_{2,1}, \ldots, \mathbf{v}_{n,1}$? Answer on page 84. That makes the last item in our list of related concepts

- The linear system represented by the augmented matrix $\begin{bmatrix} M & \mathbf{b} \end{bmatrix}$ has no solution for certain choices of $\mathbf{b}$.

Wow. Apparently there are six ways of understanding the same phenomenon.

Key Concepts

addition property of equality for all matrices A, B, C if $A = B$ then $A + C = B + C$ whenever the sums are defined.

left multiplication property of equality for all matrices A, B, C if $A = B$ then $CA = CB$ whenever the products are defined.

matrix form of a linear system if $\mathbf{v}$ is a (variable) vector with n entries, the matrix equation $M\mathbf{v} = \mathbf{b}$ is equivalent to the linear system with augmented matrix $\begin{bmatrix} M & \mathbf{b} \end{bmatrix}$ and variables $\mathbf{v}_{1,1},\ \mathbf{v}_{2,1}, \ldots, \mathbf{v}_{n,1}$.

matrix product as a linear combination of rows given matrices A and B, if BA is defined then row r of BA can be computed as a linear combination of the rows of A using row r of B as coefficients:

$$(BA)_{r,:} = B_{r,1}A_{1,:} + B_{r,2}A_{2,:} + \cdots + B_{r,n}A_{n,:}.$$

matrix product as a linear combination of columns given matrices A and B, if BA is defined then column c of BA can be computed as a linear combination of the columns of B using column c of A as coefficients:

$$(BA)_{:,c} = A_{1,c}B_{:,1} + A_{2,c}B_{:,2} + \cdots + A_{n,c}B_{:,n}.$$

In the special case of a matrix M times a vector $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}^T$,

$$M\mathbf{v} = v_1M_{:,1} + v_2M_{:,2} + \cdots + v_nM_{:,n}.$$

nontrivial solution a solution $\mathbf{v} \neq \mathbf{0}$ of the equation $M\mathbf{v} = \mathbf{0}$.

right multiplication property of equality for all matrices A, B, C if $A = B$ then $AC = BC$ whenever the products are defined.

Exercises

1. Solve

(a) $X + \begin{bmatrix} -8 & -4 \\ 19 & 6 \end{bmatrix} = \begin{bmatrix} 10 & 11 \\ 16 & 9 \end{bmatrix}$

(b) $\begin{bmatrix} -4 & -7 \\ -14 & 3 \end{bmatrix} + X = \begin{bmatrix} 12 & -20 \\ -16 & 0 \end{bmatrix}$ [𝔸]-350

(c) $\begin{bmatrix} 0 & -19 \\ -1 & 8 \end{bmatrix} - X = \begin{bmatrix} -2 & 19 \\ -14 & 20 \end{bmatrix}$ [𝕊]-294

(d) $5X + \begin{bmatrix} 1 & 0 \\ -13 & 11 \end{bmatrix} = \begin{bmatrix} -20 & -17 \\ 2 & -5 \end{bmatrix}$

2. Solve

(a) $\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} X = \begin{bmatrix} -18 & 2 \\ -2 & -8 \end{bmatrix}$

(b) $\begin{bmatrix} 5 & 2 \\ 6 & 3 \end{bmatrix} X = \begin{bmatrix} 13 & -13 \\ -19 & 7 \end{bmatrix}$ [𝕊]-294

(c) $\begin{bmatrix} 7 & 2 \\ 3 & 1 \end{bmatrix} X + \begin{bmatrix} -13 & 18 \\ -11 & 10 \end{bmatrix} = \begin{bmatrix} -20 & 14 \\ 15 & -2 \end{bmatrix}$ [𝔸]-350

(d) $\begin{bmatrix} -13 & 1 \\ 19 & 4 \end{bmatrix} - \begin{bmatrix} -5 & 9 \\ -3 & 5 \end{bmatrix} X = \begin{bmatrix} -16 & 4 \\ 1 & -6 \end{bmatrix}$

3. Solve for the specified variable. Assume all indicated operations are defined. Make a note whenever you assume a matrix is invertible.

   (a) $XYZ = B$ for Y

   (b) $XYZ = B$ for Z [𝔸]-350

   (c) $A(3B + I) = C$ for B

   (d) $PDP^{-1} = A$ for D [𝕊]-294

   (e) $2A(B^{-1} + C^T) = D$ for C [𝔸]-350

   (f) $(3C)^T + 2B^{-1} = A$ for B

4. Write the matrix equation as an equivalent linear system.

   (a) $A\mathbf{x} = \mathbf{c}$; $A = \begin{bmatrix} -6 & 2 & 19 \\ -14 & 1 & -10 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^T$; $\mathbf{c} = \begin{bmatrix} 7 & 3 \end{bmatrix}^T$

   (b) $A\mathbf{x} = \mathbf{0}$; $A = \begin{bmatrix} -34 & -3 & 37 \\ -118 & 9 & 109 \\ 26 & -3 & -23 \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} x & y & z \end{bmatrix}^T$;

   (c) $T\mathbf{r}_1 = \mathbf{r}_2$; $T = \begin{bmatrix} 10 & 14 \\ -6 & -3 \end{bmatrix}$; $\mathbf{r}_1 = \begin{bmatrix} r \\ s \end{bmatrix}$; $\mathbf{r}_2 = \begin{bmatrix} -8 \\ 9 \end{bmatrix}$ [𝔸]-350

   (d) $M\mathbf{v} = \mathbf{b}$; $M = \begin{bmatrix} -4 & -14 \\ -17 & 6 \\ -16 & -12 \end{bmatrix}$; $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 4 \\ -3 \\ 0 \end{bmatrix}$

5. Specify the matrix M, vector $\mathbf{v}$, and vector $\mathbf{b}$ so that the matrix equation $M\mathbf{v} = \mathbf{b}$ is equivalent to the linear system.

   (a) $\begin{array}{rcrcrcr} -12x & - & 6y & - & 7z & = & 16 \\ & - & 5y & + & 18z & = & 2 \\ -15x & + & 10y & + & 8z & = & -11 \end{array}$

   (b) $\begin{array}{rcrcrcr} -15x_1 & + & 4x_2 & & & = & -14 \\ -4x_1 & - & 17x_2 & - & 7x_3 & = & 3 \end{array}$

   (c) $\begin{array}{rcrcr} 15r & - & 11s & = & 0 \\ 3r & + & 10s & = & 0 \end{array}$

   (d) $\begin{array}{rcrcr} 14v_1 & - & 17v_2 & = & -11 \\ 2v_1 & & & = & -4 \\ 9v_1 & - & 2v_2 & = & -8 \end{array}$ [𝔸]-350

6. Compute the second row of the product (if it exists) without computing the rest of the product.

   (a) $\begin{bmatrix} 2 & 2 & 1 & 5 \\ 0 & 1 & -5 & -4 \\ 5 & 0 & 3 & -3 \\ 3 & -2 & -4 & 4 \\ 4 & -1 & -1 & -3 \end{bmatrix} \begin{bmatrix} -3 & 3 \\ 2 & -4 \\ -5 & 5 \\ 0 & 1 \end{bmatrix}$

   (b) $\begin{bmatrix} -3 & 2 & 3 \\ 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} 0 & 4 & -3 \\ 1 & -4 & 5 \\ 3 & 2 & -1 \end{bmatrix}$

   (c) $\begin{bmatrix} -4 & 5 & -3 & -4 \\ 0 & -3 & 3 & 2 \\ 4 & 1 & 4 & -5 \end{bmatrix} \begin{bmatrix} -3 & 3 \\ 2 & -4 \\ -5 & 5 \\ 0 & 1 \end{bmatrix}$ [𝕊]-294

   (d) $\begin{bmatrix} -3 & 2 & 3 \\ 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} 0 & 4 & -3 \\ 1 & -4 & 5 \\ 3 & 2 & -1 \end{bmatrix}$ [𝔸]-350

7. Compute the third row of the product in question 6 (if it exists) by summing an appropriate linear combination of row vectors. [𝕊]-294 [𝔸]-350

8. Compute the second column of the product in question 6 (if it exists) without computing the first. [𝕊]-295 [𝔸]-350

9. Compute the third column of the product in question 6 (if it exists) by summing an appropriate linear combination of column vectors. [𝕊]-295 [𝔸]-350

10. Suppose the third column of A contains all zeros. What can you say about the third column of BA? Why? Assume BA is defined.

11. Suppose the second row of B contains all ones. What can you say about the second row of BA? Why? Assume BA is defined.

12. Suppose the fifth column of A is three times the second column of A. What can you say about the fifth column of BA? Why? Assume BA is defined.

13. Demonstrate that the zero product rule, which holds for real numbers, does not hold for matrices. That is, show that the claim "if A and B are matrices such that $AB = 0$, then $A = 0$ or $B = 0$" is false. Do so (four times over) by providing examples of matrices A and B such that $A \neq 0$ and $B \neq 0$ yet $AB = 0$ in each of the following cases.

    (a) A and B are nonsquare matrices.

    (b) A is square but B is not square.

    (c) B is square but A is not square.

    (d) A and B are both square.

14. Argue that if $AB = 0$ then one of the following must be true.

    - $\det A = 0$
    - $\det B = 0$
    - A is not square
    - B is not square

    Use the fact that for a square matrix M, $\det M = 0$ if and only if M is noninvertible.

15. Show that the converse of (3.2.2) is true. That is, justify the claim that for all matrices A, B, C, if $A + C = B + C$ then $A = B$ whenever the sums are defined.

16. Show that the converse of (3.2.3) is false by supplying matrices A, B, C such that $CA = CB$ but $A \neq B$.

Answers

multiplication property of equality part 2 Claims (3.2.3) and (3.2.4) are distinct, and therefore both needed, because matrix multiplication is not commutative. Claim (3.2.4) can be proven as follows. Since $A = B$, $0 = A-B$. Hence, if $A = B$ then

$$\begin{aligned} 0 &= A - B \\ &= (A - B)C \\ &= AC - BC \end{aligned}$$

and therefore $AC = BC$.

$M\mathbf{v} = \mathbf{b}$ as a linear system Let M be an $m \times n$ matrix and suppose $M\mathbf{v} = \mathbf{b}$ (making $\mathbf{v}$ a vector with n entries and $\mathbf{b}$ a vector with m entries). Then

$$\begin{aligned} M\mathbf{v} &= \begin{bmatrix} M_{1,1} & M_{1,2} & \cdots & M_{1,n} \\ M_{2,1} & M_{2,2} & \cdots & M_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{m,1} & M_{m,2} & \cdots & M_{m,n} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{1,1} \\ \mathbf{v}_{2,1} \\ \vdots \\ \mathbf{v}_{n,1} \end{bmatrix} \\ &= \begin{bmatrix} M_{1,1}\mathbf{v}_{1,1} + M_{1,2}\mathbf{v}_{2,1} + \cdots + M_{1,n}\mathbf{v}_{n,1} \\ M_{2,1}\mathbf{v}_{1,1} + M_{2,2}\mathbf{v}_{2,1} + \cdots + M_{2,n}\mathbf{v}_{n,1} \\ \vdots \\ M_{m,1}\mathbf{v}_{1,1} + M_{m,2}\mathbf{v}_{2,1} + \cdots + M_{m,n}\mathbf{v}_{n,1} \end{bmatrix} \end{aligned}$$

setting this vector equal to $\mathbf{b}$ gives

$$\begin{bmatrix} M_{1,1}\mathbf{v}_{1,1} + M_{1,2}\mathbf{v}_{2,1} + \cdots + M_{1,n}\mathbf{v}_{n,1} \\ M_{2,1}\mathbf{v}_{1,1} + M_{2,2}\mathbf{v}_{2,1} + \cdots + M_{2,n}\mathbf{v}_{n,1} \\ \vdots \\ M_{m,1}\mathbf{v}_{1,1} + M_{m,2}\mathbf{v}_{2,1} + \cdots + M_{m,n}\mathbf{v}_{n,1} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_{1,1} \\ \mathbf{b}_{2,1} \\ \vdots \\ \mathbf{b}_{m,1} \end{bmatrix}$$

which can only be true if corresponding entries are equal. In other words,

$$\begin{aligned} M_{1,1}\mathbf{v}_{1,1} + M_{1,2}\mathbf{v}_{2,1} + \cdots + M_{1,n}\mathbf{v}_{n,1} &= \mathbf{b}_{1,1} \\ M_{2,1}\mathbf{v}_{1,1} + M_{2,2}\mathbf{v}_{2,1} + \cdots + M_{2,n}\mathbf{v}_{n,1} &= \mathbf{b}_{2,1} \\ &\vdots \\ M_{m,1}\mathbf{v}_{1,1} + M_{m,2}\mathbf{v}_{2,1} + \cdots + M_{m,n}\mathbf{v}_{n,1} &= \mathbf{b}_{m,1} \end{aligned}$$

Therefore, the equation $M\mathbf{v} = \mathbf{b}$ is equivalent to the system with augmented matrix $\begin{bmatrix} M & \mathbf{b} \end{bmatrix}$ and variables $\mathbf{v}_{1,1}, \mathbf{v}_{2,1}, \ldots, \mathbf{v}_{n,1}$.

3.3 Linear Independence

The matrices

$$A = \begin{bmatrix} 2 & -4 \\ 3 & -6 \end{bmatrix}, \; B = \begin{bmatrix} 4 & 2 & 6 \\ -1 & -9 & 7 \end{bmatrix}, \; C = \begin{bmatrix} 3 & 3 & -5 \\ -8 & 2 & 0 \\ 5 & -4 & 7 \end{bmatrix},$$

$$D = \begin{bmatrix} -11 & 10 & -5 & 3 & 7 & -1 \\ -8 & -1 & 6 & 6 & 0 & 1 \\ -2 & 8 & 7 & -9 & -15 & -3 \\ 8 & 5 & -10 & -7 & 5 & 2 \end{bmatrix}$$

have something in common. Each matrix has a column that can be written as a linear combination of the other columns in the matrix. Not all matrices have this property, and there is an important distinction between those that do and those that do not.

Compare $E = \begin{bmatrix} 2 & -8 \\ 1 & 5 \end{bmatrix}$ to A, for example. In E, neither column is a multiple of the other so neither column can be written as a linear combination of the other. In A, the second column is -2 times the first:

- $A_{:,2} = -2A_{:,1}$

You may be struggling a little bit to see this as a linear combination, but nothing in the definition of linear combination requires more than one term. So if an object is a multiple of another it is a linear combination of it.

Notice that $\det A = 0$ while $\det E \neq 0$. The matrix with one column that can be written as a linear combination of the others has 0 determinant while the matrix whose columns can not be written as linear combinations of the others has nonzero determinant. We made a similar observation about linear combinations of the rows of a matrix and its determinant in section 1.5.

For matrices B, C, D it is less clear that one column is a linear combination of the others, but you can check that

- $B_{:,3} = 2B_{:,1} - B_{:,2}$
- $C_{:,1} = -4C_{:,2} + 3C_{:,3}$
- $D_{:,5} = 2D_{:,1} + 2D_{:,2} + 0D_{:,3} + 3D_{:,4} + 0D_{:,6}$

Don't be misled by the suggestion that "one column" is a linear combination of the others, however. It is true, but does not tell the whole story. In no case is the column written in terms of the others special. For matrix A, for example, we could have easily pointed out that the first column is $-\frac{1}{2}$ the second. The first column is a linear combination of the second, and the second column is a linear combination of the first. Neither one should take precedence.

A little algebra will show that all the following equations are also true.

- $B_{:,2} = 2B_{:,1} - B_{:,3}$
- $B_{:,1} = \frac{1}{2}B_{:,2} + \frac{1}{2}B_{:,3}$
- $C_{:,3} = \frac{1}{3}C_{:,1} + \frac{4}{3}C_{:,2}$
- $C_{:,2} = -\frac{1}{4}C_{:,1} + \frac{3}{4}C_{:,3}$
- $D_{:,1} = -D_{:,2} - \frac{3}{2}D_{:,4} + \frac{1}{2}D_{:,5}$
- $D_{:,2} = -D_{:,1} - \frac{3}{2}D_{:,4} + \frac{1}{2}D_{:,5}$
- $D_{:,4} = -\frac{2}{3}D_{:,1} - \frac{2}{3}D_{:,2} + \frac{1}{3}D_{:,5}$

The entire set of columns involved (with nonzero coefficient) in the linear combination is special. Any such column can be written as a linear combination of the others.

To emphasize that the *set of columns* is special, not that one particular column within the matrix is special, each of the equations above can be rearranged so one side of the equation becomes 0. As a result, instead of having the 11 equations above, where in each case one column is spotlighted as the "special" column being writen in terms of the others, we have

- $2A_{:,1} + A_{:,2} = 0$
- $2B_{:,1} - B_{:,2} - B_{:,3} = 0$
- $C_{:,1} + 4C_{:,2} - 3C_{:,3} = 0$
- $2D_{:,1} + 2D_{:,2} + 0D_{:,3} + 3D_{:,4} - D_{:,5} + 0D_{:,6} = 0$

The fact that columns within the matrix can be written as linear combinations of the others is captured, but no particular column is prominent, motivating the following definition.

Let S be a set of objects on which addition and scalar mutliplication are defined and which contains an additive identity, called 0. For scalars $x_1, x_2, \dots, x_n$ and objects $b_1, b_2, \dots, b_n$ of S, we say that $b_1, b_2, \dots, b_n$ are **linearly dependent** or that $\{b_1, b_2, \dots, b_n\}$ is a linearly dependent set if there is a solution of

$$x_1 b_1 + x_2 b_2 + \cdots + x_n b_n = 0 \tag{3.3.1}$$

where not all the x_i are zero. Such a solution is called a **nontrivial solution**. Otherwise the objects $b_1, b_2, \dots, b_n$ are **linearly independent** and $\{b_1, b_2, \dots, b_n\}$ is a linearly independent set. Note that the 0 in (3.3.1) is the additive identity, not necessarily the number 0.

In addition to being a statement about the set of objects rather than one special member of the set, this definition handles the case when one of the objects in the set is the 0 object (additive identity) itself. In this case, that particular object is special. It can be written as a linear combination of the others (with all coefficients equal to zero) but it is not necessarily the case that any of the other objects can be written as a linear combination of the remaining ones. To illustrate, suppose

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

The third column is a zero vector. Accordingly,

$$E_{:,3} = 0E_{:,1} + 0E_{:,2}$$

so the third column is a linear combination of the first two. However, neither of the first two columns is a linear combination of the others. This is clear since the first two columns are not multiples of one another. In the context of the definition, we have

$$0E_{:,1} + 0E_{:,2} + E_{:,3} = 0,$$

a nontrivial linear combination (not all of the coefficients are zero) of the columns that sums to 0. No fanfare. No notes of special cases. The definition of linear dependence is clean and direct.

Matrix Characterization Part 1

Free variables, solution sets of linear systems, and pivot positions of matrices are all directly connected to the concept of linear dependence.

Theorem 5. [Characterization of Matrices Part 1] *Suppose M is an $m \times n$ matrix, $\mathbf{v}$ has n entries, and $\mathbf{b}$ has m entries. Then the following are equivalent.*

(i) The columns of M are linearly independent.

(ii) No column of M is a linear combination of the others.

(iii) $M\mathbf{v} = \mathbf{0}$ has only the trivial solution.

(iv) M has a pivot position in every column.

(v) There is a matrix L such that $LM = I$.

(vi) $M\mathbf{v} = \mathbf{b}$ has at most one solution for each $\mathbf{b}$.

(vii) $M\mathbf{v} = \mathbf{b}$ has no free variables.

The following list of arguments will show that if one of the statements is true, so is another...and if that one is true so is a third...and if that one is true so is the next...and so on until all the statements have been justified. Such a series of justifications means that if the first statement is true, they are all true since they all followed logically from the first. Proving they are equivalent requires one more step. The last statement will be shown to imply the first, completing a logical path from any one of the statements to any other. Closing the loop this way means that if *any one* of the statements is true, they are all true, the very meaning of **equivalent**!

Crumpet 18: Proof by Contraposition

Suppose you are trying to prove that if some statement, call it p, is true, then some other statement, call it q is true. In short, you are trying to prove that if p is true then q is true. Then it is just as good to prove the contrapositive claim, if q is false then p is false, because if the contrapositive is true then it is impossible to have both q false and p true. In other words, if p is true so is q because q cannot be false at the same time p is true (and that means if p is true then q is true). By similar logic, if any statement in a list of equivalent statements is false, they are all false.

The following arguments demonstrate that (i) ⇒ (ii), (ii) ⇒ (iii), (iii) ⇒ (iv), (iv) ⇒ (v), (v) ⇒ (vi), (vi) ⇒ (vii), (vii) ⇒ (iii), and (iii) ⇒ (i). More succinctly, (i) ⇒ (ii) ⇒ (iii) ⇒ (iv) ⇒ (v) ⇒ (vi) ⇒ (vii) ⇒ (iii) ⇒ (i), and diagrammatically,

$$\begin{array}{ccccccc} \text{(i)} & & & & \text{(iv)} & \Rightarrow & \text{(v)} \\ \Downarrow & \nwarrow & & \nearrow & & & \Downarrow \\ \text{(ii)} & \Rightarrow & \text{(iii)} & \Leftarrow & \text{(vii)} & \Leftarrow & \text{(vi)} \end{array}$$

The diagram illustrates a logical path from any one of the statements to any other. Therefore, justifying each statement as claimed shows that the statements are equivalent.

(i) ⇒ (ii) Requested in exercise 20.

(ii) ⇒ (iii) Suppose $M\mathbf{v} = \mathbf{0}$ has a nontrivial solution, $\mathbf{v} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$. Then $M\mathbf{v} = x_1M_{:,1} + x_2M_{:,2} + \cdots + x_nM_{:,n} = \mathbf{0}$, and since $\mathbf{v}$ is a nontrivial solution, one of the entries of $\mathbf{v}$ is nonzero, say x_i. Therefore, $x_iM_{:,i} = -x_1M_{:,1} - \ldots - x_{i-1}M_{:,i-1} - x_{i+1}M_{:,i+1} - \ldots - x_nM_{:,n}$ and more to the point,

$$M_{:,i} = -\frac{x_1}{x_i}M_{:,1} - \ldots - \frac{x_{i-1}}{x_i}M_{:,i-1} - \frac{x_{i+1}}{x_i}M_{:,i+1} - \ldots - \frac{x_n}{x_i}M_{:,n}$$

making column i a linear combination of the other columns.

(iii) ⇒ (iv) Suppose M does not have a pivot position in every column. Then $M\mathbf{v} = \mathbf{0}$, a consistent system with solution $\mathbf{v} = \mathbf{0}$, has free variables. By theorem 1, $M\mathbf{v} = \mathbf{0}$ has more than one solution.

(iv) ⇒ (v) Let R be the reduced row echelon form of M. Because M has a pivot position in every column the first m rows of R, $R_{1:m,:}$, must be the $m \times m$ identity matrix. Because $R = E_k \cdots E_2E_1M$ for some $n \times n$ elementary matrices $E_1, E_2, \ldots, E_k$, we have $R = EM$ where $E = E_k \cdots E_2E_1$. Hence $E_{1:m,:}M = R_{1:m,:} = I$. Let $L = E_{1:m,:}$.

(v) ⇒ (vi) Suppose $LM = I$ and $M\mathbf{v} = \mathbf{b}$. Then $L(M\mathbf{v}) = L\mathbf{b} \Rightarrow (LM)\mathbf{v} = L\mathbf{b} \Rightarrow I\mathbf{v} = L\mathbf{b} \Rightarrow \mathbf{v} = L\mathbf{b}$. Hence $M\mathbf{v} = \mathbf{b}$ has exactly one solution, $\mathbf{v} = L\mathbf{b}$. So $M\mathbf{v} = \mathbf{b}$ either has zero or one (in other words, at most one) solution.

(vi) ⇒ (vii) Suppose $M\mathbf{v} = \mathbf{b}$ has a free variable (for every $\mathbf{b}$) and let $\mathbf{b} = M_{:,1}$. Then $\mathbf{v} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}^T$ is a solution of $M\mathbf{v} = \mathbf{b}$ so $M\mathbf{v} = \mathbf{b}$ is consistent. By theorem 1 a consistent linear system with free variables has infinitely many solutions.

(vii) ⇒ (iii) Suppose $M\mathbf{v} = \mathbf{b}$ has no free variables (for any $\mathbf{b}$). Then $M\mathbf{v} = \mathbf{0}$ has no free variables and is consistent, having $\mathbf{v} = \mathbf{0}$ as a solution. By theorem 1, $M\mathbf{v} = \mathbf{0}$ has exactly one solution, the trivial solution.

(iii) ⇒ (i) Let $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}^T$. By assumption, $M\mathbf{v} = v_1M_{:,1} + v_2M_{:,2} + \cdots + v_nM_{:,n} = \mathbf{0}$ has only the trivial solution. By definition of linear independence, $M_{:,1}, M_{:,2}, \cdots, M_{:,n}$ (the columns of M) are linearly independent.

Later, we will see that the determinant, row equivalence, invertibility, and function concepts are also directly connected to these statements.

Key Concepts

characterization of matrices see theorem 5.

equivalent statements a list of statements such that if one of them is true they are all true.

linearly dependent objects $b_1, b_2, \ldots, b_n$ are linearly dependent whenever $\{b_1, b_2, \ldots, b_n\}$ is a linearly dependent set.

linearly independent objects $b_1, b_2, \ldots, b_n$ are linearly independent whenever $\{b_1, b_2, \ldots, b_n\}$ is a linearly independent set.

linearly independent set a set that is not linearly dependent.

linearly dependent set a set for which a nontrivial linear combination of its elements equals the additive identity.

nontrivial linear combination a linear combination in which not all the coefficients are zero.

nontrivial solution any solution $x_1, x_2, \ldots, x_n$ of the equation $x_1b_1 + x_2b_2 + \cdots + x_nb_n = 0$ where not all the x_i are zero.

trivial solution the solution $x_1 = x_2 = \cdots = x_n = 0$ of the equation $x_1b_1 + x_2b_2 + \cdots + x_nb_n = 0$.

Exercises

1. Show that the vectors are linearly independent.

   (a) $\begin{bmatrix} 5 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ [𝕊]-295

   (b) $\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \end{bmatrix}$

   (c) $\begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 \\ -3 \\ -3 \end{bmatrix}$ [𝕊]-296

   (d) $\begin{bmatrix} 3 \\ 2 \\ -4 \end{bmatrix}, \begin{bmatrix} -2 \\ -4 \\ 3 \end{bmatrix}$

   (e) $\begin{bmatrix} -5 \\ -4 \\ -1 \end{bmatrix}, \begin{bmatrix} -5 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ -2 \end{bmatrix}$

   (f) $\begin{bmatrix} -2 \\ 5 \\ 1 \end{bmatrix}, \begin{bmatrix} -5 \\ -4 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}$

   (g) $\begin{bmatrix} 1 \\ 0 \\ -1 \\ -2 \end{bmatrix}, \begin{bmatrix} -4 \\ -2 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} -3 \\ -2 \\ 3 \\ 3 \end{bmatrix}$

   (h) $\begin{bmatrix} -1 \\ -3 \\ 5 \\ -4 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \\ -3 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ -5 \\ 5 \\ -1 \end{bmatrix}$

2. Show that the linear system has at most one solution for any values b_1, b_2, b_3, b_4.

   (a) $\begin{array}{rcrcl} 4x & + & 11y & = & b_1 \\ 5x & + & 12y & = & b_2 \end{array}$

   (b) $\begin{array}{rcrcl} -x_1 & + & x_2 & = & b_1 \\ -8x_1 & + & 7x_2 & = & b_2 \\ 5x_1 & + & x_2 & = & b_3 \end{array}$

   (c) $\begin{array}{rcrcrcl} 2v_1 & - & v_2 & - & v_3 & = & b_1 \\ 7v_1 & + & 2v_2 & & & = & b_2 \\ -4v_1 & - & v_2 & + & v_3 & = & b_3 \end{array}$ [𝕊]-296

   (d) $\begin{array}{rcrcrcl} x & + & 5y & + & 7z & = & b_1 \\ 6x & + & 7y & + & 2z & = & b_2 \\ -6x & + & 3y & - & z & = & b_3 \\ 5x & + & y & & & = & b_4 \end{array}$ [𝕊]-297

   (e) $\begin{array}{rcrcl} -3x & + & y & = & b_1 \\ 11x & + & 4y & = & b_2 \end{array}$

   (f) $\begin{array}{rcrcl} -x_1 & & & = & b_1 \\ 8x_1 & - & 7x_2 & = & b_2 \\ 6x_1 & + & 5x_2 & = & b_3 \end{array}$

   (g) $\begin{array}{rcrcrcl} -2v_1 & + & 4v_2 & - & 6v_3 & = & b_1 \\ 3v_1 & + & 6v_2 & + & 7v_3 & = & b_2 \\ 6v_1 & - & 8v_2 & + & 2v_3 & = & b_3 \end{array}$

   (h) $\begin{array}{rcrcrcl} x & + & 7y & + & z & = & b_1 \\ 8x & + & 8y & + & 6z & = & b_2 \\ -4x & + & 6y & - & 7z & = & b_3 \\ -7x & + & 2y & - & 5z & = & b_4 \end{array}$

3. Show that the homogeneous system has only one solution, the trivial solution.

   (a) $\begin{bmatrix} 1 & 8 \\ 1 & -5 \end{bmatrix} \mathbf{v} = \mathbf{0}$ [𝕊]-297

   (b) $\begin{bmatrix} 6 & 5 \\ 5 & -2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

   (c) $\begin{bmatrix} -2 & 5 \\ 5 & -8 \\ -2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

   (d) $\begin{bmatrix} 5 & 1 \\ -3 & 3 \\ 2 & -2 \end{bmatrix} \mathbf{x} = \mathbf{0}$

   (e) $\begin{bmatrix} 5 & 6 & 1 \\ 6 & -7 & -2 \\ -1 & -4 & -1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

(f) $\begin{bmatrix} -4 & -2 & -1 \\ -3 & 3 & 4 \\ 2 & 1 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{0}$

(g) $\begin{bmatrix} 6 & 3 & -1 \\ 5 & 0 & 1 \\ 1 & -4 & 1 \\ 5 & 7 & -4 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \mathbf{0}$ [𝕊]-297

(h) $\begin{bmatrix} 0 & 1 & 4 \\ -4 & 3 & 5 \\ 1 & 1 & -1 \\ -3 & -1 & 5 \end{bmatrix}\mathbf{v} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

4. Show that the functions are linearly independent.
   (a) $1 + t$, $t + t^2$, and $1 + t^2$
   (b) $\sin^2 t$ and $\cos^2 t$ [𝕊]-297
   (c) e^x and e^{2x}

5. Show that the functions are linearly dependent.
   (a) $1 + 3t - 2t^2$, $-9 - 23t + 21t^2$, and $1 + 7t + t^2$
   (b) 1, $\sin^2 t$, and $\cos^2 t$
   (c) $\sin^2 t$, $\cos^2 t$, and $\cos(2t)$ [𝕊]-298

6. Do the columns of the matrix form a linearly independent set?

(a) $\begin{bmatrix} 12 & 21 \\ -24 & 58 \end{bmatrix}$

(b) $\begin{bmatrix} -3 & 9 \\ 11 & 1 \\ -6 & 7 \end{bmatrix}$

(c) $\begin{bmatrix} -8 & -5 & -7 \\ 5 & -9 & -6 \end{bmatrix}$

(d) $\begin{bmatrix} -1 & -5 & -10 \\ 1 & -8 & 6 \\ -9 & -4 & 10 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & -1 & 5 & 4 \\ -2 & -11 & 9 & -7 \\ 6 & -5 & 11 & -6 \end{bmatrix}$

(f) $\begin{bmatrix} 2 & -2 & -9 \\ 0 & -11 & -8 \\ 9 & 6 & -5 \\ 1 & -7 & 8 \end{bmatrix}$

(g) $\begin{bmatrix} -54 & -30 & -96 & 9 & 6 & -74 \\ -24 & 4 & -93 & -68 & 21 & 15 \\ 70 & -89 & 78 & 26 & -78 & 0 \\ -46 & 68 & -87 & -88 & -39 & 67 \end{bmatrix}$ [𝕊]-298

7. Determine whether the set is linearly independent.

(a) $\{-10t + t^2 - 5t^3, 2t + 6t^2 - 2t^3, 31t + 14t^3\}$

(b) $\left\{\begin{bmatrix} 1 & 8 & -11 \end{bmatrix}, \begin{bmatrix} 9 & 4 & -7 \end{bmatrix}, \begin{bmatrix} 4 & -2 & 2 \end{bmatrix}\right\}$ [𝕊]-298

(c) $\left\{\begin{bmatrix} 3 & -4 \\ -6 & -3 \end{bmatrix}, \begin{bmatrix} 21 & 6 \\ -18 & 13 \end{bmatrix}, \begin{bmatrix} 6 & 9 \\ 0 & 11 \end{bmatrix}\right\}$

8. Is the empty set linearly independent or linearly dependent? [𝔸]-350

9. A 9×6 matrix has linearly independent columns. How many pivot positions does it have?

10. A 7×6 matrix has 5 pivot positions. What can you say about the linear independence of its columns?

11. Give an example of a 3×2 matrix M such that $M\mathbf{v} = \mathbf{0}$
    (a) has only the trivial solution
    (b) has a nontrivial solution

12. What are the possible reduced row echelon forms of a 4×3 matrix with [𝔸]-350
    (a) linearly independent columns?
    (b) linearly dependent columns?

13. What are the possible row echelon forms of a 2×2 matrix with
    (a) linearly independent columns?
    (b) linearly dependent columns?

14. If M is an $m \times n$ matrix with linearly independent columns, what can you say about the relationship between m and n? HINT: Can a matrix with linearly independent columns have more columns than rows?

15. Find the value(s) of x for which the matrix has linearly independent columns.

(a) $\begin{bmatrix} 2 & -6 \\ 3 & x \end{bmatrix}$

(b) $\begin{bmatrix} x & 4 \\ -2 & 7 \end{bmatrix}$ [𝔸]-350

(c) $\begin{bmatrix} 1 & 8 & 0 \\ 6 & x & 1 \\ -3 & -2 & 5 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 8 & 2 \\ 6 & 45 & 1 \\ -3 & -20 & x \end{bmatrix}$ [𝔸]-350

16. Find the value(s) of x for which the matrix has linearly dependent columns.

(a) $\begin{bmatrix} 3 & x \\ -5 & 4 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 7 \\ x & 5 \end{bmatrix}$ [𝔸]-350

(c) $\begin{bmatrix} x & -6 & 27 \\ -2 & 8 & -30 \\ -1 & 5 & -18 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 5 & -4 \\ -5 & 3 & x \\ -7 & -11 & 4 \end{bmatrix}$ [𝔸]-350

17. Find a nontrivial solution of $M\mathbf{v} = \mathbf{0}$ using the fact that the first and second columns of M are identical. Do not use row reduction.

$$M = \begin{bmatrix} 94 & 94 & 85 \\ -97 & -97 & 83 \\ 6 & 6 & 24 \\ 5 & 5 & -77 \end{bmatrix}$$

18. Argue that the statements are equivalent. [𝕊]-298

    (a) $x = 8$

    (b) x is a perfect cube between 6 and 20.

19. Argue that the statements are equivalent.

    (a) The graph of f is a line with slope 3 and y-intercept -5.

    (b) $f(x) = 3x - 5$.

    (c) f is a first degree polynomial passing through $(-10, -35)$ and $(10, 25)$.

    HINT: This requires three separate arguments.

For exercises 20 and 21 assume that addition and scalar multiplication are defined on the objects and that there exists an additive identity.

20. Argue that if a set of objects is linearly independent then none of the objects is a linear combination of the others. HINT: Try proof by contraposition with multiple cases. Suppose one of the objects in the set is a linear combination of the others, and logically conclude that the set is linearly dependent.

21. Argue that if none of the objects of a set is a linear combination of the others, then the set is linearly independent. HINT: Try proof by contraposition. Suppose the set is linearly dependent, and logically conclude that one of the objects in the set is a linear combination of the others.

The key ingredient in the proof of the uniqueness of reduced row echelon form (crumpet 23 on page 160) is that *row operations do not affect the linear dependence relationships among the columns of a matrix*. To illustrate the claim, three matrices are given in exercises 22 and 23—M; F, a row echelon form of M; and R, the reduced row echelon form of M.

22. Linear dependence is maintained. (i) Find a nontrivial linear combination of (some of) the columns of R that sums to $\mathbf{0}$. (ii) Check that the same linear combination of columns of F sums to $\mathbf{0}$. (iii) Check that the same linear combination of columns of M sums to $\mathbf{0}$.

    (a) $$R = \begin{bmatrix} 1 & 0 & -11/8 & 0 \\ 0 & 1 & 9/8 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
    $$F = \begin{bmatrix} 9 & 3 & -9 & 4 \\ 0 & -8 & -9 & 10 \\ 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
    $$M = \begin{bmatrix} 0 & 120 & 135 & -136 \\ -45 & -279 & -252 & 289 \\ 0 & 56 & 63 & -63 \\ 9 & 243 & 261 & -268 \end{bmatrix}$$

    (b) $$R = \begin{bmatrix} 1 & -7/9 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
    $$F = \begin{bmatrix} -9 & 7 & -11 & -11 \\ 0 & 0 & 10 & -3 \\ 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
    $$M = \begin{bmatrix} 180 & -140 & 950 & -20 \\ -108 & 84 & -512 & -11 \\ -72 & 56 & -358 & 0 \\ -189 & 147 & -971 & 12 \end{bmatrix}$$ [𝔸]-350

23. Linear independence is maintained. (i) Find two different pairs of linearly independent columns of R. (ii) Check that the same two pairs of columns of F are linearly independent. (iii) Check that the same two pairs of columns of M are linearly independent.

    (a) $$R = \begin{bmatrix} 1 & -6/11 & 0 & 3/11 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
    $$F = \begin{bmatrix} 11 & -6 & -2 & -5 \\ 0 & 0 & 3 & 12 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
    $$M = \begin{bmatrix} -33 & 18 & 18 & 63 \\ -99 & 54 & 21 & 57 \\ 1056 & -576 & -204 & -528 \\ 253 & -138 & -49 & -127 \end{bmatrix}$$

    (b) $$R = \begin{bmatrix} 1 & -1 & 0 & -89/56 \\ 0 & 0 & 1 & -1/7 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
    $$F = \begin{bmatrix} -8 & 8 & 12 & 11 \\ 0 & 0 & 7 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
    $$M = \begin{bmatrix} -64 & 64 & 166 & 78 \\ -264 & 264 & 543 & 342 \\ -392 & 392 & 826 & 505 \\ -136 & 136 & 288 & 175 \end{bmatrix}$$ [𝔸]-350

3.4 Characterization of $m \times n$ Matrices

Theorem 5 of section 3.3 has a counterpart phrased in terms of the rows of M. Parts (i), (ii), (iii), and (vii) of the following theorem can be justified through reference to the previous, but parts (iv), (v), and (vi) cannot.

Theorem 6. [Characterization of Matrices Part 2] *Suppose M is an $m \times n$ matrix, $\mathbf{v}$ and $\mathbf{c}$ have n entries, and $\mathbf{b}$ and $\mathbf{w}$ have m entries. Then the following are equivalent.*

(i) The rows of M are linearly independent.

(ii) No row of M is a linear combination of the others.

(iii) $\mathbf{w}^T M = \mathbf{0}^T$ has only the trivial solution.

(iv) M has a pivot position in every row.

(v) There is a matrix R such that $MR = I$.

(vi) $M\mathbf{v} = \mathbf{b}$ has at least one solution for every $\mathbf{b}$.

(vii) $\mathbf{w}^T M = \mathbf{c}^T$ has no free variables.

The justification for this theorem proceeds by logically connecting the statements according to the following diagram.

$$\begin{array}{ccccc} & & \text{(vii)} & & \\ & \swarrow & & \nwarrow & \\ \text{(i)} & \Rightarrow & \text{(ii)} & \Rightarrow & \text{(iii)} \\ \Downarrow & & & & \Uparrow \\ \text{(iv)} & \Rightarrow & \text{(vi)} & \Rightarrow & \text{(v)} \end{array}$$

Though the first several implications can be proven without reference to theorem 5, such reference will be used to emphasize the direct connection between the two theorems.

(i) ⇒ (ii) The rows of M are linearly independent, so the columns of M^T are linearly independent. By theorem 5, none of the columns of M^T can be written as a linear combination of the others, so none of the rows of M can be writen as a linear combination of the others.

(ii) ⇒ (iii) No row of M can be written as a linear combination of the others, so no column of M^T can be written as a linear combination of the others. By theorem 5, $M^T\mathbf{w} = \mathbf{0}$ has only the trivial solution. Since $M^T\mathbf{w} = \mathbf{0}$ is equivalent to $(M^T\mathbf{w})^T = \mathbf{0}^T$ (transpose both sides), which is equivalent to $\mathbf{w}^T M = \mathbf{0}^T$ (simplifying the lefthand side), the conclusion follows.

(iii) ⇒ (vii) Since $\mathbf{w}^T M = \mathbf{0}^T$ has only the trivial solution, the equivalent equations $(\mathbf{w}^T M)^T = (\mathbf{0}^T)^T$ and $M^T\mathbf{w} = \mathbf{0}$ have only the trivial solution. By theorem 5, $M^T\mathbf{w} = \mathbf{c}$ has no free variables. Therefore, the equivalent equation $\mathbf{w}^T M = \mathbf{c}^T$ has no free variables.

(vii) ⇒ (i) Since $\mathbf{w}^T M = \mathbf{c}^T$ has no free variables, the equivalent equation $M^T\mathbf{w} = \mathbf{c}$ has no free variables. By theorem 5, the columns of M^T are linearly independent. Therefore, the rows of M are linearly independent.

(i) ⇒ (iv) Suppose M does not have a pivot in every row. Then any row echelon form of M has a row of zeros. Since that row of zeros is the result of a nontrivial linear combination of the rows of M, there is a nontrivial linear combination of the rows of M that sum to $\mathbf{0}^T$. Therefore the rows of M are not linearly independent.

(iv) ⇒ (vi) Since M has a pivot in every row, no row echelon form of M has a row of zeros. Therefore $\begin{bmatrix} M & \mathbf{b} \end{bmatrix}$ cannot have a pivot in the rightmost column, and by theorem 1 the system $M\mathbf{v} = \mathbf{b}$ is consistent (has at least one solution) for any $\mathbf{b}$.

(vi) ⇒ (v) Let $\mathbf{b}_j = (I_{m\times m})_{:,j}$ for $j = 1, 2, \ldots, m$. Since $M\mathbf{v} = \mathbf{b}$ has a solution for every $\mathbf{b}$, there is a vector $\mathbf{v}_j$ such that $M\mathbf{v}_j = \mathbf{b}_j$. Letting $R = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \end{bmatrix}$ we have $MR = I$.

(v) ⇒ (iii) Suppose there is a matrix R such that $MR = I$. If $\mathbf{w}^T M = \mathbf{0}^T$, the following equations are deduced by matrix algebra.

$$\begin{aligned}(\mathbf{w}^T M)R &= \mathbf{0}^T R\\ \mathbf{w}^T(MR) &= \mathbf{0}^T\\ \mathbf{w}^T I &= \mathbf{0}^T\\ \mathbf{w}^T &= \mathbf{0}^T\end{aligned}$$

Hence $\mathbf{w}^T = \mathbf{0}^T$ ($\mathbf{w} = \mathbf{0}$) is the only solution of $\mathbf{w}^T M = \mathbf{0}^T$.

The justification for many of the upcoming claims relies heavily on induction (see axiom 5 of crumpet 16 on page 72). The principle behind induction is to show that (i) the claim is actually true for some particular integer, and (ii) if the claim is true for some integer at least as large, then it is also true for the successive integer. This way, part (i) establishes the claim for a particular integer, say k. Then part (ii) establishes the claim for the successive integer, $k+1$. Part (ii) also establishes the claim for the successor to the successive integer, $k+2$. Applying part (ii) again establishes the claim for the $k+3$, and so on, part (ii) establishing the claim for all integers greater than k. Induction is often the most practical way to show that a statement is true for all integers and is particularly useful in proving claims about matrices of size n (for all n). Proofs of this nature will be shown, but this is not a course on proof technique, so it is up to you or your instructor to decide how deeply you need to understand these proofs. Even if you are not prepared to write your own induction proof or fully understand one, reading them is a good way to get a feel for the technique.

An **upper triangular matrix** is one in which all entries below the main diagonal are zero, and a **lower triangular matrix** is one in which all entries above the main diagonal are zero. Using $\star$ to represent any number (as in the notation of section 2.3), a square upper triangular matrix looks like

$$\begin{bmatrix}\star & \star & \star & \star & \cdots & \star\\ 0 & \star & \star & \star & \cdots & \star\\ 0 & 0 & \star & \star & \cdots & \star\\ 0 & 0 & 0 & \star & \cdots & \star\\ \vdots & \vdots & \vdots & \vdots & \ddots & \star\\ 0 & 0 & 0 & 0 & \cdots & \star\end{bmatrix}$$

and a square lower triangular matrix looks like

$$\begin{bmatrix}\star & 0 & 0 & 0 & \cdots & 0\\ \star & \star & 0 & 0 & \cdots & 0\\ \star & \star & \star & 0 & \cdots & 0\\ \star & \star & \star & \star & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \ddots & 0\\ \star & \star & \star & \star & \cdots & \star\end{bmatrix}.$$

All the nonzero entries are above or on the main diagonal (the upper triangle) for an upper triangular matrix, and all the nonzero entries are below or on the main diagonal (the lower triangle) for a lower triangular matrix.

The determinant of a lower triangular matrix is the product of the entries on its diagonal. Now is a good time to work out a couple examples on your own and think about why this is true in general. We will prove it by induction, but the proof may not resonate with you the way your own thoughts about it will.

Claim. If L is a lower triangular $n \times n$ matrix, then $\det L = L_{1,1}L_{2,2}\cdots L_{n,n}$.

Proof. If L is a 1×1 matrix, it is upper triangular and $\det L = \det([L_{1,1}]) = L_{1,1}$. This establishes part (i) of the proof. The claim is true for the particular value $n = 1$. Now we assume that the claim is true for some (arbitrary) value $n = k$ greater than or equal to one. That is, if L is a lower triangular $k \times k$ matrix and $k \geq 1$, then $\det L = L_{1,1}L_{2,2}\cdots L_{k,k}$. To complete the proof, we must use this information to prove that if L is a $(k+1)\times(k+1)$ matrix, the next size up, then $\det L = L_{1,1}L_{2,2}\cdots L_{k+1,k+1}$. To that end suppose L is a $(k+1)\times(k+1)$ matrix. By definition,

$$\det L = (-1)^{1+1}L_{1,1}\det L_{\backslash 1,1} + (-1)^{1+2}L_{1,2}\det L_{\backslash 1,2}\cdots + (-1)^{1+3}L_{1,3}\det L_{\backslash 1,3} \tag{3.4.1}$$

Since L is lower triangular, $L_{1,j} = 0$ whenever $j > 1$. Therefore, all the terms of the determinant are zero except the first one. The determinant simplifies to

$$\det L = L_{1,1} \det L_{\backslash 1,1}. \tag{3.4.2}$$

But $L_{\backslash 1,1}$ is a $k \times k$ matrix, so its determinant is the product of the entries on its diagonal (this is our inductive hypothesis). So $\det L_{\backslash 1,1} = L_{2,2}L_{3,3}\cdots L_{k+1,k+1}$. Substituting this expression into (3.4.2), we have $\det L = L_{1,1}L_{2,2}L_{3,3}\cdots L_{k+1,k+1}$, and the proof is complete. □

Exercises 17 through 19 request a proof that the determinant of a square upper triangular matrix is also the product of the entries on the main diagonal.

Key Concepts

characterization of matrices see theorem 6.

upper triangular matrix a matrix in which all entries below the main diagonal are zero.

lower triangular matrix a matrix in which all entries above the main diagonal are zero.

determinant of a (square) lower triangular matrix the product of the entries on the main diagonal.

determinant of a (square) upper triangular matrix the product of the entries on the main diagonal.

proof by induction showing that (i) the claim is true for some particular integer, and (ii) if the claim is true for some integer at least as large, then it is also true for the next integer. These together prove that the claim is true for all integers greater than or equal to the particular integer of part (i).

Exercises

1. The size and number of pivot positions of a matrix M are given. Answer the following questions as completely as you can. (i) Are the rows of M linearly independent? (ii) Are the columns of M linearly independent? (iii) How many solutions does $M\mathbf{v} = \mathbf{0}$ have? (iv) How many solutions does $M\mathbf{v} = \mathbf{b}$ have for arbitrary $\mathbf{b}$?
   (a) 5×8; 5
   (b) 3×3; 2
   (c) 7×7; 7
   (d) 9×6; 6 [𝔸]-351

2. The size of a matrix M is given. (i) What is the maximum number of pivot positions M could have? Assume it has that maximum number and answer the following questions. (ii) Are the rows of M linearly independent? (iii) Are the columns of M linearly independent? (iv) How many solutions does $M\mathbf{v} = \mathbf{0}$ have? (v) How many solutions does $M\mathbf{v} = \mathbf{b}$ have for arbitrary $\mathbf{b}$?
   (a) 13×5
   (b) 12×12
   (c) 9×29 [𝔸]-351

3. Redo question 2 parts (ii)-(v) assuming M has less than the maximum number of pivot positions. [𝔸]-351

4. Show that the linear system has at least one solution for any values b_1, b_2, b_3.
   (a) $$\begin{array}{rcrcl} 4x &+& 11y &=& b_1 \\ 5x &+& 12y &=& b_2 \end{array}$$
   (b) $$\begin{array}{rcrcrcl} -x_1 &-& 8x_2 &+& 5x_3 &=& b_1 \\ x_1 &+& 7x_2 &+& x_3 &=& b_2 \end{array}$$
   (c) $$\begin{array}{rcrcrcl} 2v_1 &-& v_2 &-& v_3 &=& b_1 \\ 7v_1 &+& 2v_2 && &=& b_2 \\ -5v_1 &-& v_2 &+& v_3 &=& b_3 \end{array}$$
   (d) $$\begin{array}{rcrcrcrcl} w &+& 6x &-& 6y &+& 5z &=& b_1 \\ 5w &+& 7x &+& 3y &+& z &=& b_2 \\ 7w &+& 2x &-& y && &=& b_3 \end{array}$$

5. Show that the homogeneous system has only one solution, the trivial solution.
   (a) $\mathbf{v}^T \begin{bmatrix} 1 & 8 \\ 1 & -5 \end{bmatrix} = \mathbf{0}^T$
   (b) $\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} -2 & 5 & -2 \\ 5 & -8 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$
   (c) $\mathbf{x}^T \begin{bmatrix} 5 & 6 & 1 \\ 6 & -7 & -2 \\ -1 & -4 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$
   (d) $\begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} \begin{bmatrix} 6 & 5 & 1 & 5 \\ 3 & 0 & -4 & 7 \\ -1 & 1 & 1 & -4 \end{bmatrix} = \mathbf{0}^T$ [𝕊]-299

6. Do the rows of the matrix form a linearly independent set?
   (a) $\begin{bmatrix} 12 & 21 \\ -16 & -28 \end{bmatrix}$
   (b) $\begin{bmatrix} -3 & 6 \\ 4 & 11 \\ -6 & 7 \end{bmatrix}$

(c) $\begin{bmatrix} -3 & 6 & -12 \\ -2 & 2 & -11 \end{bmatrix}$

(d) $\begin{bmatrix} -1 & -3 & -10 \\ -7 & 9 & 5 \\ 2 & -5 & 0 \end{bmatrix}$

(e) $\begin{bmatrix} -18 & 1 & -1 & 7 \\ -6 & 2 & 6 & 3 \\ 12 & -5 & 0 & 4 \end{bmatrix}$ [𝕊]-299

(f) $\begin{bmatrix} 5 & 2 & -3 \\ -4 & -6 & 7 \\ 12 & 9 & -5 \\ 0 & -12 & -11 \end{bmatrix}$

(g) $\begin{bmatrix} -38 & -5 & 30 & 25 & -44 \\ 4 & -28 & 44 & -39 & 43 \\ 21 & 42 & 1 & -11 & -29 \\ 47 & 12 & 26 & -10 & -8 \\ -13 & -15 & 39 & -9 & 22 \\ 9 & -3 & 3 & -41 & 49 \end{bmatrix}$ [𝕊]-299

7. A 5×8 matrix has linearly independent rows. How many pivot positions does it have? [𝔸]-351
8. A 6×7 matrix has 5 pivot positions. What can you say about the linear indpendence of its rows?
9. Give an example of a 2×3 matrix M such that $\mathbf{v}^T M = \mathbf{0}^T$
   (a) has only the trivial solution
   (b) has a nontrivial solution
10. What are the possible reduced row echelon forms of a 3×4 matrix with [𝔸]-351
    (a) linearly independent rows?
    (b) linearly dependent rows?
11. What are the possible row echelon forms of a 2×2 matrix with
    (a) linearly independent rows?
    (b) linearly dependent rows?

    Compare your answer with the answer to section 3.3 exercise 13.
12. Find the determinant.

(a) $\begin{bmatrix} 2 & 0 \\ -137 & -3 \end{bmatrix}$

(b) $\begin{bmatrix} -8 & 0 & 0 \\ -15 & \frac{1}{4} & 0 \\ -29 & -1 & -41 \end{bmatrix}$ [𝔸]-351

(c) $\begin{bmatrix} -4 & 0 & 0 & 0 \\ -15 & -\frac{5}{12} & 0 & 0 \\ 32 & -4 & 3 & 0 \\ 27 & 37 & 41 & 5 \end{bmatrix}$

13. If M is an $m \times n$ matrix with linearly independent rows, what can you say about the relationship between m and n? HINT: Can a matrix with linearly independent rows have more rows than columns?
14. Find the value(s) of x for which the matrix has linearly independent rows.

(a) $\begin{bmatrix} 2 & -6 \\ 3 & x \end{bmatrix}$

(b) $\begin{bmatrix} x & 4 \\ -2 & 7 \end{bmatrix}$ [𝔸]-351

(c) $\begin{bmatrix} 1 & 8 & 0 \\ 6 & x & 1 \\ -3 & -2 & 5 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 8 & 2 \\ 6 & 45 & 1 \\ -3 & -20 & x \end{bmatrix}$ [𝔸]-351

Compare your answer with the answer to section 3.3 exercise 15.

15. Find the value(s) of x for which the matrix has linearly dependent rows.

(a) $\begin{bmatrix} 3 & x \\ -5 & 4 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 7 \\ x & 5 \end{bmatrix}$ [𝔸]-351

(c) $\begin{bmatrix} x & -6 & 27 \\ -2 & 8 & -30 \\ -1 & 5 & -18 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 5 & -4 \\ -5 & 3 & x \\ -7 & -11 & 4 \end{bmatrix}$ [𝔸]-351

Compare your answer with the answer to section 3.3 exercise 16.

16. Find a nontrivial solution of $\mathbf{v}^T M = \mathbf{0}^T$ using the fact that the first and second rows of M are identical. Do not use row operations.

$$M = \begin{bmatrix} -14 & -29 & 49 & -32 \\ -14 & -29 & 49 & -32 \\ 44 & -25 & 13 & -35 \end{bmatrix}$$

17. Prove that if U is upper triangular, so is $U_{\backslash 1,j}$. [𝕊]-299
18. Prove that if U is upper triangular, then $U_{\backslash 1,j}$ has a zero on its main diagonal whenever $j > 1$. [𝕊]-299
19. Prove that if U is an upper triangular $n \times n$ matrix, then $\det U = U_{1,1}U_{2,2}\cdots U_{n,n}$. HINT: Use the facts proven in exercises 17 and 18. [𝕊]-299

3.5 The Determinant Revisited

Pick a number. Any number.

Add 6.

Multiply (your new number) by 6.

Subtract 9.

Divide by 3.

Subtract 9.

$\vdots$

Tell me your latest number, and I'll tell you your starting number. It's half of your latest number! Putting the instructions into symbols, you are being asked to calculate $\frac{6(x+6)-9}{3} - 9$, which simplifies to $2x$. Hence the number you start with, x, will always be half of what you end with!

Matrices and row operations work in a similar fashion, and you can be the magician. I'll pick a matrix...any matrix...and tell you what it looks like after some operation. After swapping the first two rows, my matrix is

$$\begin{bmatrix} 23 & 12 & 22 \\ 13 & -11 & -19 \\ -16 & 19 & -5 \end{bmatrix}.$$

What was my original matrix? Answer on page 105.

How about another? After scaling the third row of my matrix by 2, my matrix is

$$\begin{bmatrix} 2 & -13 & -6 \\ -7 & -9 & 21 \\ -14 & -18 & 24 \end{bmatrix}.$$

What was my original matrix? Answer on page 105.

And one last one...after replacing the first row of my matrix by the first row plus three times the third, my matrix is

$$\begin{bmatrix} 11 & 9 & -1 \\ 3 & 5 & -8 \\ 3 & 8 & 1 \end{bmatrix}.$$

What was my original matrix? Answer on page 105.

In each case the row operation can be undone to recover the original matrix. This is exactly the concept of an inverse! The six elementary matrices corresponding to the six row operations (the row operations that gave the matrices above and the row operations used to recover the original matrices), in the order encountered are

$$\begin{array}{cc} \text{operation} & \text{recovery} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \\ \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

It must therefore be that each operation matrix is invertible and

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

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

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

Multiplying the operation matrix by the inverse matrix will verify the inverse pairs. Each elementary matrix has an inverse elementary matrix of the same type, and in the case of a row swap, the elementary matrix is its own inverse.

Now notice a couple of things about the determinants:

$$\begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} = -1$$

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{vmatrix} = 2 \qquad \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \end{vmatrix} = \frac{1}{2}$$

$$\begin{vmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1 \qquad \begin{vmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1$$

Come to think of it, the determinant of a scale matrix will always be the scale factor! Can you justify this claim? Answer on page 105. Wait a minute! the determinant of a replace matrix is always 1. Can you justify this claim? Answer on page 105.

Is the determinant of a swap matrix always -1? There is no such thing as a swap matrix with one row. There is only one swap matrix with two rows:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and there are only three swap matrices with three rows:

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

It is easy enough to check that all four of these matrices have determinant -1, so maybe all swap matrices do have determinant -1.

Crumpet 19: Binomial Coefficients

The number of swap matrices with n rows, $n > 1$, is $\frac{n(n-1)}{2}$. For example, there are $\frac{4(3)}{2} = 6$ swap matrices with 4 rows, and there are $\frac{100(99)}{2} = 4950$ swap matrices with 100 rows. The formula $\frac{n(n-1)}{2}$ is a special case of the "choose formula", which is a formula for the number of ways to choose k objects from a set of n objects, with $0 \leq k \leq n$. This number is also known as a binomial coefficient and there are several notations for it. Common notations and the formula for "n choose k" are

$$\binom{n}{k} = {}_nC_k = C(n,k) = \frac{n!}{k!(n-k)!}.$$

Proving it for all $n \times n$ swap matrices requires induction, but before we can do it cleanly, we need one more fact: if M is an $n \times n$ matrix with $n > 1$, then

$$\begin{aligned}\det M &= (-1)^{i+1} M_{i,1} \det M_{\setminus i,1} + (-1)^{i+2} M_{i,2} \det M_{\setminus i,2} + \cdots + (-1)^{i+n} M_{i,n} \det M_{\setminus i,n} \\ &= (-1)^{1+j} M_{1,j} \det M_{\setminus 1,j} + (-1)^{2+j} M_{2,j} \det M_{\setminus 2,j} + \cdots + (-1)^{n+j} M_{n,j} \det M_{\setminus n,j}. \end{aligned} \tag{3.5.1}$$

for any i from 1 through n or any j from 1 through n. This formula implies that determinants may be computed by expanding along any row or any column, not just the first row. To illustrate,

expansion along row 1

$$\begin{aligned} &\begin{vmatrix} -5 & 4 & -1 \\ 2 & -2 & 3 \\ 5 & -2 & -3 \end{vmatrix} \\ &= -5\begin{vmatrix} -2 & 3 \\ -2 & -3 \end{vmatrix} - 4\begin{vmatrix} 2 & 3 \\ 5 & -3 \end{vmatrix} - 1\begin{vmatrix} 2 & -2 \\ 5 & -2 \end{vmatrix} \\ &= -5(6+6) - 4(-6-15) - 1(-4+10) \\ &= -60 + 84 - 6 \\ &= 18 \end{aligned}$$

expansion along row 3

$$\begin{aligned} &\begin{vmatrix} -5 & 4 & -1 \\ 2 & -2 & 3 \\ 5 & -2 & -3 \end{vmatrix} \\ &= 5\begin{vmatrix} 4 & -1 \\ -2 & 3 \end{vmatrix} + 2\begin{vmatrix} -5 & -1 \\ 2 & 3 \end{vmatrix} - 3\begin{vmatrix} -5 & 4 \\ 2 & -2 \end{vmatrix} \\ &= 5(12-2) + 2(-15+2) - 3(10-8) \\ &= 50 - 26 - 6 \\ &= 18 \end{aligned}$$

expansion along column 2

$$\begin{aligned} &\begin{vmatrix} -5 & 4 & -1 \\ 2 & -2 & 3 \\ 5 & -2 & -3 \end{vmatrix} \\ &= -4\begin{vmatrix} 2 & 3 \\ 5 & -3 \end{vmatrix} - 2\begin{vmatrix} -5 & -1 \\ 5 & -3 \end{vmatrix} + 2\begin{vmatrix} -5 & -1 \\ 2 & 3 \end{vmatrix} \\ &= -4(-6-15) - 2(15+5) + 2(-15+2) \\ &= 84 - 40 - 26 \\ &= 18 \end{aligned}$$

Formula (3.5.1) makes it relatively straightforward to prove that all swap matrices have determinant -1. We have already shown (assuming you did the calculation above) that the determinant of any 2×2 matrix (and there is only one of them) is -1. Proceeding by induction, assume that for some $k \geq 2$, the determinant of all $k \times k$ swap matrices is -1, and let S be a particular but arbitrary $(k+1) \times (k+1)$ swap matrix where rows i and j have been swapped. Since $k + 1 \geq 3$, there must be a row of S that is not involved in the swap, say row ℓ. Then expanding the determinant of S along row ℓ yields

$$\begin{aligned} \det S &= (-1)^{\ell+1} S_{\ell,1} \det S_{\setminus \ell,1} + (-1)^{\ell+2} S_{\ell,2} \det S_{\setminus \ell,2} + \cdots + (-1)^{\ell+k+1} S_{\ell,k+1} \det S_{\setminus \ell,k+1} \\ &= (-1)^{\ell+\ell} S_{\ell,\ell} \det S_{\setminus \ell,\ell} \end{aligned}$$

because $S_{\ell,:}$ is the ℓ^{th} row of the identity matrix (not being involved in the swap), meaning $S_{\ell,m} = 0$ whenever $m \neq \ell$. Since the swapped rows are both in $S_{\setminus \ell,\ell}$, $S_{\setminus \ell,\ell}$ is a $k \times k$ swap matrix and the inductive hypothesis implies $\det S_{\setminus \ell,\ell} = -1$. Therefore,

$$\det S = (-1)^{\ell+\ell} S_{\ell,\ell} \det S_{\setminus \ell,\ell} = (1)(1)(-1) = -1,$$

completing the proof.

Crumpet 20: Proof of Formula (3.5.1)

Proving formula (3.5.1) requires a bit of work. One way to do it is to prove that (i) there is only one function G taking $n \times n$ matrices as inputs and returning scalars as outputs with the following four properties and (ii) the expressions in formula (3.5.1) have these four properties. Thus, each expression must give the same result.

1. $G(I) = 1$
2. $G(A) = 0$ whenever A has two identical columns.

3. If A, B, and C are identical except in their k^{th} columns where $C_{:,k} = A_{:,k} + B_{:,k}$, then $G(C) = G(A) + G(B)$.
4. If A and B are identical except in their k^{th} columns where $A_{:,k} = cB_{:,k}$, then $G(A) = cG(B)$.

To begin, suppose G is a function from $n \times n$ matrices to scalars satisfying the four properties above. Then G also has the following two properties.

5. $G(A) = 0$ whenever the columns of A are linearly dependent. *Proof:* Because the columns of A are linearly dependent, one of them, say column k, can be written as a linear combination of the others. That is, $A_{:,k} = \sum_{j \neq k} c_j A_{:,j}$ for some constants c_j. Then

$$\begin{aligned} G(A) &= G\left(\begin{bmatrix} A_{:,1} & \cdots & A_{:,k-1} & \sum_{j\neq k} c_j A_{:,j} & A_{:,k+1} & \cdots & A_{:,n} \end{bmatrix}\right) \\ &= \sum_{j\neq k} G\left(\begin{bmatrix} A_{:,1} & \cdots & A_{:,k-1} & c_j A_{:,j} & A_{:,k+1} & \cdots & A_{:,n} \end{bmatrix}\right) \\ &= \sum_{j\neq k} c_j G\left(\begin{bmatrix} A_{:,1} & \cdots & A_{:,k-1} & A_{:,j} & A_{:,k+1} & \cdots & A_{:,n} \end{bmatrix}\right) \\ &= \sum_{j\neq k} c_j \cdot 0 = 0 \end{aligned}$$

   by applying properties 3, 4, and 2, respectively.

6. $G(B) = -G(A)$ whenever B is the result of swapping two columns of A. *Proof:* Suppose B is the result of swapping columns i and j of A, and without loss of generality, assume $i < j$. Then

$$B = \begin{bmatrix} A_{:,1} & \cdots & A_{:,j} & \cdots & A_{:,i} & \cdots & A_{:,n} \end{bmatrix}.$$

   Now let $C = \begin{bmatrix} A_{:,1} & \cdots & A_{:,i} + A_{:,j} & \cdots & A_{:,i} + A_{:,j} & \cdots & A_{:,n} \end{bmatrix}$. Then by repeated application of property 3,

$$\begin{aligned} G(C) = G(A) + G(B) \\ + G\left(\begin{bmatrix} A_{:,1} & \cdots & A_{:,i} & \cdots & A_{:,i} & \cdots & A_{:,n} \end{bmatrix}\right) \\ + G\left(\begin{bmatrix} A_{:,1} & \cdots & A_{:,j} & \cdots & A_{:,j} & \cdots & A_{:,n} \end{bmatrix}\right) \end{aligned}$$

   and by property 2 the last two terms are zero as is $G(C)$. Hence, $0 = G(A) + G(B)$, concluding the proof.

To begin the induction proof that G is unique, note that $G([a]) = G(a[1]) = aG([1]) = aG(I) = a$ by properties 4 and 1, so G is uniquely determined for 1×1 matrices. Now suppose G is unique for all $(k-1) \times (k-1)$ matrices for some $k > 1$, and let M be a particular but arbitrary $k \times k$ matrix. If the columns of M are linearly dependent, then property 5 implies $G(M) = 0$, so $G(M)$ is uniquely determined. Now suppose the columns of M are linearly independent. By theorem 5, M has a pivot in every column. Since M is square, M has a pivot in every row. Therefore M has a nonzero entry in row k, say $M_{k,j} \neq 0$. Letting

$$A = \begin{bmatrix} M_{:,1} - \frac{M_{k,1}}{M_{k,j}} M_{:,j} & \cdots & M_{:,j-1} - \frac{M_{k,j-1}}{M_{k,j}} M_{:,j} & M_{:,j+1} - \frac{M_{k,j+1}}{M_{k,j}} M_{:,j} & \cdots & M_{:,n} - \frac{M_{k,k}}{M_{k,j}} M_{:,j} & M_{:j} \end{bmatrix},$$

repeated application of properties 3 and 4 plus property 6 if needed implies $G(A) = \pm G(M)$ depending on whether $j = k$. Either way, $G(M)$ is uniquely determined if $G(A)$ is. Note that $A_{k,:} = \begin{bmatrix} 0 & \cdots & 0 & M_{k,j} \end{bmatrix}$, making it sufficient to show that $H(B)$ defined on $(k-1) \times (k-1)$ matrices by

$$H(B) = \frac{1}{M_{k,j}} G\left(\begin{bmatrix} B & M_{1:k-1,j} \\ \mathbf{0}^T & M_{k,j} \end{bmatrix}\right)$$

is uniquely determined. But $H(B)$ inherits properties 2, 3, and 4 from G, so it only remains to establish that $H(I) = 1$.

Then, by the inductive hypothesis, H is uniquely determined. To that end,

$$\begin{aligned} H(I) &= \frac{1}{M_{k,j}} G\left(\begin{bmatrix} I & M_{1:k-1,j} \\ \mathbf{0}^T & M_{k,j} \end{bmatrix}\right) \\ &= \frac{1}{M_{k,j}} G\left(\begin{bmatrix} I_{:,1} & \cdots & I_{:,k-1} & \sum_{i=1}^{k} M_{i,j} I_{:,i} \end{bmatrix}\right) \\ &= \frac{1}{M_{k,j}} \sum_{i=1}^{k} M_{i,j} G\left(\begin{bmatrix} I_{:,1} & \cdots & I_{:,k-1} & I_{:,i} \end{bmatrix}\right) \\ &= \frac{1}{M_{k,j}} \left(M_{k,j} G(I)\right) \\ &= 1 \end{aligned}$$

concluding the proof that G is unique.

To complete the proof of formula (3.5.1), it remains to show that each of the two expressions for $\det M$ has properties 1 through 4. This is because formula (1.5.1), the definition of determinant, is one of the expressions, so by uniqueness they all produce the same result (as the determinant).

Proceeding by induction, note that $\det([a]) = a$ satisfies all four properties, so the determinant has all four properties on 1×1 matrices. Now suppose $\det M$ has all four properties on $(\ell - 1) \times (\ell - 1)$ matrices for some $\ell > 1$ and consider, for any fixed $i = 1, 2, \ldots \ell$, the formula

$$\det M = (-1)^{i+1} M_{i,1} \det M_{\backslash i,1} + (-1)^{i+2} M_{i,2} \det M_{\backslash i,2} + \cdots + (-1)^{i+\ell-1} M_{i,\ell} \det M_{\backslash i,\ell}$$

on $\ell \times \ell$ matrices.

1. $$\begin{aligned} \det I_{\ell\times\ell} &= (-1)^{i+1} I_{i,1} \det I_{\backslash i,1} + (-1)^{i+2} I_{i,2} \det I_{\backslash i,2} + \cdots + (-1)^{i+\ell} I_{i,\ell} \det I_{\backslash i,\ell} \\ &= (-1)^{i+i} I_{i,i} \det I_{\backslash i,i} \end{aligned}$$

   since $I_{i,j} = 0$ whenever $j \neq i$. But $I_{\backslash i,i} = I_{(\ell-1)\times(\ell-1)}$, so by the inductive hypothesis $\det I_{\backslash i,i} = 1$ and we have $\det I_{\ell\times\ell} = (-1)^{2i} I_{i,i}(1) = (1)(1)(1) = 1$.

2. Suppose M is a particular but arbitrary $\ell\times\ell$ matrix with columns j and k identical, and without loss of generality assume $j < k$. Then

   $$\begin{aligned} \det M &= (-1)^{i+1} M_{i,1} \det M_{\backslash i,1} + (-1)^{i+2} M_{i,2} \det M_{\backslash i,2} + \cdots + (-1)^{i+\ell} M_{i,\ell} \det M_{\backslash i,\ell} \\ &= (-1)^{i+j} M_{i,j} \det M_{\backslash i,j} + (-1)^{i+k} M_{i,k} \det M_{\backslash i,k} \end{aligned}$$

   since $M_{\backslash i,m}$ has two identical columns whenever $m \notin \{j, k\}$ and therefore $\det M_{\backslash i,m} = 0$ by the inductive hypothesis. But because columns j and k of M are identical, $M_{i,j} = M_{i,k}$, so we can rewrite $\det M = (-1)^{i+j} M_{i,j} \left[\det M_{\backslash i,j} + (-1)^{k-j} \det M_{\backslash i,k}\right]$. Now if we swap columns j and $j+1$ of $M_{\backslash i,k}$, and then columns $j+1$ and $j+2$, and so on to column k, a total of $k - j - 1$ swaps, the result is $M_{\backslash i,j}$, so by the inductive hypothesis $\det M_{\backslash i,j} = (-1)^{k-j-1} \det M_{\backslash i,k}$. Substituting into the latest expression for $\det M$, we have $\det M = (-1)^{i+j} M_{i,j} \left[(-1)^{k-j-1} \det M_{\backslash i,k} + (-1)^{k-j} \det M_{\backslash i,k}\right] = (-1)^{i+k-1} \left[\det M_{\backslash i,k} - \det M_{\backslash i,k}\right] = 0$.

3. Suppose A, B, and C are identical $\ell \times \ell$ matrices except in their k^{th} columns where $C_{:,k} = A_{:,k} + B_{:,k}$. Observe that $C_{\backslash i,k} = B_{\backslash i,k} = A_{\backslash i,k}$, $C_{i,j} = B_{i,j} = A_{i,j}$, and $\det C_{\backslash i,j} = \det A_{\backslash i,j} + \det B_{\backslash i,j}$ (by the inductive hypothesis) for

$j \neq k$ and all i. It then follows that

$$\begin{aligned}
\det C &= (-1)^{i+1} C_{i,1} \det C_{\backslash i,1} + (-1)^{i+2} C_{i,2} \det C_{\backslash i,2} + \cdots + (-1)^{i+\ell} C_{i,\ell} \det C_{\backslash i,\ell} \\
&= (-1)^{i+k} C_{i,k} \det C_{\backslash i,k} + \sum_{j \neq k} (-1)^{i+j} C_{i,j} \det C_{\backslash i,j} \\
&= (-1)^{i+k} \left(A_{i,k} + B_{i,k}\right) \det C_{\backslash i,k} + \sum_{j \neq k} (-1)^{i+j} C_{i,j} \left[\det A_{\backslash i,j} + \det B_{\backslash i,j}\right] \\
& (-1)^{i+k} A_{i,k} \det C_{\backslash i,k} + (-1)^{i+k} B_{i,k} \det C_{\backslash i,k} \\
&\quad + \sum_{j \neq k} (-1)^{i+j} \left[C_{i,j} \det A_{\backslash i,j} + C_{i,j} \det B_{\backslash i,j}\right] \\
&= (-1)^{i+k} A_{i,k} \det A_{\backslash i,k} + (-1)^{i+k} B_{i,k} \det B_{\backslash i,k} \\
&\quad + \sum_{j \neq k} (-1)^{i+j} \left[A_{i,j} \det A_{\backslash i,j} + B_{i,j} \det B_{\backslash i,j}\right] \\
&= \sum_{j=1}^{\ell} (-1)^{i+j} \left[A_{i,j} \det A_{\backslash i,j} + B_{i,j} \det B_{\backslash i,j}\right] = \det A + \det B
\end{aligned}$$

4. Suppose A and B are identical $\ell \times \ell$ matrices except in their k^{th} columns where $A_{:,k} = cB_{:,k}$. Observe that $B_{\backslash i,k} = A_{\backslash i,k}$, $B_{i,j} = A_{i,j}$, and $\det A_{\backslash i,j} = c \det B_{\backslash i,j}$ (by the inductive hypothesis) for $j \neq k$ and all i. It then follows that

$$\begin{aligned}
\det A &= (-1)^{i+1} A_{i,1} \det A_{\backslash i,1} + (-1)^{i+2} A_{i,2} \det A_{\backslash i,2} + \cdots + (-1)^{i+\ell} A_{i,\ell} \det A_{\backslash i,\ell} \\
&= (-1)^{i+k} A_{i,k} \det A_{\backslash i,k} + \sum_{j \neq k} (-1)^{i+j} A_{i,j} \det A_{\backslash i,j} \\
&= (-1)^{i+k} c B_{i,k} \det B_{\backslash i,k} + \sum_{j \neq k} (-1)^{i+j} B_{i,j} \left(c \det B_{\backslash i,j}\right) \\
&= c \left[(-1)^{i+k} B_{i,k} \det B_{\backslash i,k} + \sum_{j \neq k} (-1)^{i+j} B_{i,j} \det B_{\backslash i,j}\right] \\
&= c \sum_{j=1}^{\ell} (-1)^{i+j} B_{i,j} \det B_{\backslash i,j} = c \det B
\end{aligned}$$

Hence the determinant may be calculated by expansion along any row.

As for expansion along any column, we begin by showing that the function $f(M) = \det M^T$ (where $\det M$ is defined by row expansion) has all four properties for any size matrix, so must equal $\det M$. Note that if M is a 1×1 matrix, $M^T = M$. Therefore $f(M) = \det M^T = \det M = M_{1,1}$, so $f(M)$ has all four properties. Observe that if M is an $n \times n$ matrix and $n > 1$, then by definition of f and the row expansion formula for determinant,

$$f(M) = \sum_{j=1}^{n} (-1)^{i+j} M_{j,i} \det(M^T)_{\backslash i,j} \tag{3.5.2}$$

for any fixed $i = 1, 2, \ldots n$. We now proceed to show that f has all four properties for $n \times n$ matrices where $n > 1$.

1. For any n, $I_{n\times n}^T = I_{n\times n}$, so $f(I_{n\times n}) = \det(I_{n\times n}^T) = \det(I_{n\times n}) = 1$.

2. Suppose M is a 2×2 matrix with two identical columns. Then $M = \begin{bmatrix} a & a \\ b & b \end{bmatrix}$ for some constants a, b, and

$$f(M) = \det M^T = \det\left(\begin{bmatrix} a & b \\ a & b \end{bmatrix}\right) = ab - ba = 0.$$

Now set $n > 2$, suppose $f(A) = 0$ for any $(n-1) \times (n-1)$ matrix A with two identical columns, and let M be an $n \times n$ matrix with identical columns k and ℓ. Because M has at least three columns, there is an i, $1 \leq i \leq n$

such that $i \neq k$ and $i \neq \ell$. Then, for this particular i,

$$\begin{aligned} f(M) &= \sum_{j=1}^{n} (-1)^{i+j} M_{j,i} \det(M^T)_{\backslash i,j} \\ &= \sum_{j=1}^{n} (-1)^{i+j} M_{j,i} \det(M_{\backslash j,i})^T \\ &= \sum_{j=1}^{n} (-1)^{i+j} M_{j,i} f(M_{\backslash j,i}) \end{aligned}$$

By the inductive hypothesis, $f(M_{\backslash j,i}) = 0$ for all j, so $f(M) = 0$.

3. Let $n > 1$ and suppose A, B, and C are identical $n \times n$ matrices except in their k^{th} columns where $C_{:,k} = A_{:,k} + B_{:,k}$. Then $(C^T)_{\backslash k,j} = (A^T)_{\backslash k,j} = (B^T)_{\backslash k,j}$ for all $j = 1, \ldots, n$ because A^T, B^T, and C^T are all identical except in their k^{th} rows. Now, applying (100) with $i = k$,

$$\begin{aligned} f(C) &= \sum_{j=1}^{n} (-1)^{k+j} C_{j,k} \det(C^T)_{\backslash k,j} \\ &= \sum_{j=1}^{n} (-1)^{k+j} \left(A_{j,k} + B_{j,k}\right) \det(C^T)_{\backslash k,j} \\ &= \sum_{j=1}^{n} (-1)^{k+j} \left(A_{j,k} \det(C^T)_{\backslash k,j} + B_{j,k} \det(C^T)_{\backslash k,j}\right) \\ &= \sum_{j=1}^{n} (-1)^{k+j} \left(A_{j,k} \det(A^T)_{\backslash k,j} + B_{j,k} \det(B^T)_{\backslash k,j}\right) \\ &= \sum_{j=1}^{n} (-1)^{k+j} A_{j,k} \det(A^T)_{\backslash k,j} + \sum_{j=1}^{\ell} (-1)^{k+j} B_{j,k} \det(B^T)_{\backslash k,j} \\ &= f(A) + f(B) \end{aligned}$$

4. Let $n > 1$ and suppose A and B are identical $n \times n$ matrices except in their k^{th} columns where $A_{:,k} = cB_{:,k}$. Observe that $(A^T)_{\backslash k,j} = (B^T)_{\backslash k,j}$ for all $j = 1, \ldots, n$ because A^T and B^T are identical except in their k^{th} rows. Now applying (100) with $i = k$,

$$\begin{aligned} f(A) &= \sum_{j=1}^{n} (-1)^{k+j} A_{j,k} \det(A^T)_{\backslash k,j} \\ &= \sum_{j=1}^{n} (-1)^{k+j} c B_{j,k} \det(B^T)_{\backslash k,j} \\ &= c \sum_{j=1}^{n} (-1)^{k+j} B_{j,k} \det(B^T)_{\backslash k,j} \\ &= c f(B) \end{aligned}$$

Finally, the expression

$$(-1)^{1+j} M_{1,j} \det M_{\backslash 1,j} + (-1)^{2+j} M_{2,j} \det M_{\backslash 2,j} + \cdots + (-1)^{n+j} M_{n,j} \det M_{\backslash n,j}$$

from (3.5.1) equals

$$\begin{aligned} \sum_{i=1}^{n} (-1)^{i+j} M_{i,j} \det M_{\backslash i,j} &= \sum_{i=1}^{n} (-1)^{j+i} M^T_{j,i} \det\left(M^T_{\backslash j,i}\right)^T \\ &= \sum_{i=1}^{n} (-1)^{j+i} M^T_{j,i} \det M^T_{\backslash j,i} \\ &= \det M^T \\ &= \det M \end{aligned}$$

by the row expansion formula and the fact that $\det M^T = \det M$.

This proof is an adaptation of the presentation in sections 6.1 and 6.2 of [17].

To recap,

- the determinant of a swap matrix is -1,
- the determinant of a scale matrix is the scale factor, and
- the determinant of a replace matrix is 1.

Interestingly, within the proof of (3.5.1) lie the proofs that

- if B is the result of swapping two columns of a square matrix A, then $\det B = -\det A$, and
- if B is the result of scaling a column of a square matrix A by c, then $\det B = c \det A$, and
- for any square matrix M, $\det M^T = \det M$.

Putting these three facts together, it is easy to justify the following two facts.

- If B is the result of swapping two rows of a square matrix A, then $\det B = -\det A$.

  [$\det B = \det B^T = \det A^T = -\det A$ since B^T is the result of swapping two columns of A^T and $\det M^T = \det M$.]
- If B is the result of scaling a row of a square matrix A by c, then $\det B = c \det A$.

  Can you justify this? Answer on page 105.

The relevance of all these observations is mounting evidence that $\det(EA) = \det E \cdot \det A$ for any square matrix A and elementary matrix E. This will be an important point soon enough. We already have that $\det(EA) = -\det A = \det E \cdot \det A$ when E is a swap matrix and $\det(EA) = c \det A = \det E \cdot \det A$ when E is a scale matrix. We are only missing this fact for elementary replacement matrices.

The proof of (3.5.1) does not provide direct proof that if B is the result of a row replacement in a square matrix A, then $\det B = \det A$, but it provides the right tools for the job. Beside the facts already noted, we learn from the proof that

- if A, B, and C are identical square matrices except in one column, say the k^{th}, where $C_{:,k} = A_{:,k} + B_{:,k}$, then $\det(C) = \det A + \det B$, and
- if C is a square matrix with two identical columns, then $\det C = 0$.

As we just encountered, statements about the columns of a matrix and its determinant can generally be restated in terms of rows since $\det M^T = \det M$. It is safe to conclude that

- if A, B, and C are identical square matrices except in one row, say the k^{th}, where $C_{k,:} = A_{k,:} + B_{k,:}$, then $\det(C) = \det A + \det B$, and
- if C is a square matrix with two identical rows, then $\det C = 0$.

Justifications are requested in exercises 15 and 16. These two facts plus the fact that if B is the result of scaling a row of a square matrix A by c, then $\det B = c \det A$ make it a straighforward matter to prove that if B is the result of a row replacement in a square matrix A, then $\det B = \det A$.

Proof. Let A be an $n \times n$ matrix and suppose B is the result of adding c times the j^{th}row to the k^{th} row of A, $j \neq k$. Then

$$\det B = \begin{vmatrix} A_{1,:} \\ \vdots \\ A_{k-1,:} \\ A_{k,:} + cA_{j,:} \\ A_{k+1} \\ \vdots \\ A_{n,:} \end{vmatrix} = \begin{vmatrix} A_{1,:} \\ \vdots \\ A_{k-1,:} \\ A_{k,:} \\ A_{k+1} \\ \vdots \\ A_{n,:} \end{vmatrix} + \begin{vmatrix} A_{1,:} \\ \vdots \\ A_{k-1,:} \\ cA_{j,:} \\ A_{k+1} \\ \vdots \\ A_{n,:} \end{vmatrix} = \det A + c \begin{vmatrix} A_{1,:} \\ \vdots \\ A_{k-1,:} \\ A_{j,:} \\ A_{k+1} \\ \vdots \\ A_{n,:} \end{vmatrix} = \det A + 0 = \det A.$$

□

Therefore, $\det(EA) = 1 \det A = \det E \cdot \det A$ when E is a replacement matrix.

Key Concepts

elementary matrices are invertible and $\det(EA) = \det E \cdot \det A$ for any square matrix A and elementary matrix E.

determinant by expansion the determinant of an $n \times n$ matrix M may be calculated by expansion along any row or any column:

$$\begin{aligned} \det M &= (-1)^{i+1} M_{i,1} \det M_{\setminus i,1} + (-1)^{i+2} M_{i,2} \det M_{\setminus i,2} + \cdots + (-1)^{i+n} M_{i,n} \det M_{\setminus i,n} \\ &= (-1)^{1+j} M_{1,j} \det M_{\setminus 1,j} + (-1)^{2+j} M_{2,j} \det M_{\setminus 2,j} + \cdots + (-1)^{n+j} M_{n,j} \det M_{\setminus n,j} \end{aligned}$$

for any $i = 1, \ldots, n$ or any $j = 1, \ldots, n$.

determinant of replacement matrix if E is an elementary replacement matrix, $\det E = 1$.

determinant of swap matrix if E is an elementary swap matrix, $\det E = -1$.

deteminant of scale matrix if E is an elementary scale matrix, $\det E = s$ where s is the scale factor.

determinant of the transpose for any square matrix A, $\det A^T = \det A$.

Exercises

1. Take advantage of the fact that the determinant may be expanded along any row or any column to compute the determinant.

(a) $\begin{bmatrix} 0 & -2 & 0 & 9 \\ -4 & 0 & 1 & 0 \\ 4 & -9 & 0 & -2 \\ 0 & 0 & 0 & 4 \end{bmatrix}$ [𝕊]-300

(b) $\begin{bmatrix} 0 & -5 & 0 & -2 \\ 0 & 7 & 0 & 0 \\ 4 & -3 & 0 & 0 \\ 2 & 0 & 2 & -3 \end{bmatrix}$

(c) $\begin{bmatrix} 3 & 0 & 6 & 0 \\ 0 & 5 & 0 & 3 \\ 5 & 0 & -6 & -3 \\ 0 & -9 & 0 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & 9 & 3 & -2 \\ 0 & -2 & 0 & 0 \\ -6 & -4 & 0 & 0 \\ 0 & 4 & 2 & 0 \end{bmatrix}$

2. Find the determinant of the triangular matrix.

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

(b) $\begin{bmatrix} 3 & 2 & 6 \\ 0 & 8 & -1 \\ 0 & 0 & 5 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & 8 & 0 & 0 \\ 1 & 8 & 4 & 0 \\ -2 & -2 & 8 & 7 \end{bmatrix}$ [𝕊]-300

(d) $\begin{bmatrix} -2 & 7 & 7 & 6 \\ 0 & 1 & 3 & -1 \\ 0 & 0 & -1 & 6 \\ 0 & 0 & 0 & -1 \end{bmatrix}$

3. Use the fact that

$$\begin{vmatrix} -2 & 0 & 0 \\ 8 & -2 & 2 \\ 8 & 0 & 8 \end{vmatrix} = 32$$

to compute the determinant of [𝕊]-300

(a) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -2 & 0 & 0 \\ 8 & -2 & 2 \\ 8 & 0 & 8 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{128} \end{bmatrix}\begin{bmatrix} -2 & 0 & 0 \\ 8 & -2 & 2 \\ 8 & 0 & 8 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{7}{9} \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -2 & 0 & 0 \\ 8 & -2 & 2 \\ 8 & 0 & 8 \end{bmatrix}$

(d) $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 18 & 0 & 1 \end{bmatrix}\begin{bmatrix} -2 & 0 & 0 \\ 8 & -2 & 2 \\ 8 & 0 & 8 \end{bmatrix}$

4. Use the fact that

$$\begin{vmatrix} 5 & 3 & 7 \\ -1 & 4 & 3 \\ 2 & 6 & 8 \end{vmatrix} = 14$$

to compute the determinant of

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

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

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \sqrt{17} & 1 \end{bmatrix}\begin{bmatrix} 5 & 3 & 7 \\ -1 & 4 & 3 \\ 2 & 6 & 8 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 5 & 3 & 7 \\ -1 & 4 & 3 \\ 2 & 6 & 8 \end{bmatrix}$

5. Use the fact that

$$\begin{vmatrix} 4 & 4 & -1 \\ 1 & 7 & 3 \\ -1 & 3 & 4 \end{vmatrix} = 38$$

to compute the determinant of [A]-351

(a) $\begin{bmatrix} 4 & 4 & -1 \\ -1 & 3 & 4 \\ 1 & 7 & 3 \end{bmatrix}$

(b) $\begin{bmatrix} 4 & 4 & -1 \\ 1 & 7 & 3 \\ -\frac{1}{2} & \frac{3}{2} & 2 \end{bmatrix}$

(c) $\begin{bmatrix} 4 & 4 & -1 \\ 0 & 10 & 7 \\ -1 & 3 & 4 \end{bmatrix}$

(d) $\begin{bmatrix} 5 & 35 & 15 \\ 4 & 4 & -1 \\ -1 & 3 & 4 \end{bmatrix}$

6. Use the fact that

$$\begin{vmatrix} 1 & 6 & 5 \\ 1 & -1 & 8 \\ 7 & 0 & 5 \end{vmatrix} = 336$$

to compute the determinant of

(a) $\begin{bmatrix} 1 & 6 & 5 \\ 1 & -1 & 8 \\ \frac{1}{2} & 0 & \frac{5}{14} \end{bmatrix}$

(b) $\begin{bmatrix} 7 & 0 & 5 \\ 1 & -1 & 8 \\ 1 & 6 & 5 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 6 & 5 \\ 1 & -1 & 8 \\ 9 & -2 & 21 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & -1 & 8 \\ \frac{1}{84} & \frac{1}{14} & \frac{5}{84} \\ 7 & 0 & 5 \end{bmatrix}$

7. Use the fact that

$$\begin{vmatrix} 6 & -2 & 1 \\ 6 & 5 & -2 \\ 7 & 7 & 8 \end{vmatrix} = 455$$

to compute the determinant of

(a) $\begin{bmatrix} 6 & 6 & 7 \\ -2 & 5 & 7 \\ 1 & -2 & 8 \end{bmatrix}$

(b) $\begin{bmatrix} 6 & 7 & 6 \\ -2 & 7 & 5 \\ 1 & 8 & -2 \end{bmatrix}$

(c) $\begin{bmatrix} 6 & 6 & 7 \\ -2 & 5 & 7 \\ 1 & -2 & 8 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ \pi & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(d) $\begin{bmatrix} 6 & 6 & 7 \\ -2 & 5 & 7 \\ 1 & -2 & 8 \end{bmatrix}\begin{bmatrix} \frac{1}{35} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

8. Use the facts that

$$\det(E_3E_2E_1A) = 1$$

and

(a) E_1 is a swap matrix.

(b) E_2 is a scale matrix with scale factor -2.

(c) E_3 is a replacement matrix.

to determine $\det A$. [A]-351

9. Use the facts that

$$\det(E_4E_3E_2E_1A) = 1$$

and

(a) E_1 is a scale matrix with scale factor 2.

(b) E_2 is a scale matrix with scale factor 3.

(c) E_3 is a replacement matrix.

(d) E_4 is a scale matrix with scale factor $\frac{1}{36}$.

to determine $\det A$.

10. Let A be a 4×4 matrix with $\det A = 3$. Find $\det(2A)$.

11. Let M be an $n \times n$ matrix with $\det M = \frac{1}{3}$. Find $\det(7M)$. [A]-351

12. Let A be a square matrix and E a scale matrix with scale factor $\frac{2}{3}$. Find $\det(E^3A)$. That is, $\det(EEEA)$.

13. Suppose A is a square matrix, E is a swap matrix, and $\det(EA) = 33$. Find

    (a) $\det A$

    (b) $\det E$

    (c) $\det A^T$

14. Suppose M is a square matrix, E is a replacement matrix, and $\det(EM) = -\frac{1}{2}$. Find $\det(M^T)$. [𝔸]-351

15. Use the fact that if A, B, and C are identical square matrices except in one column, the k^{th}, where $C_{:,k} = A_{:,k} + B_{:,k}$, then $\det(C) = \det A + \det B$ to prove that *if A, B, and C are identical square matrices except in one row, say the k^{th}, where $C_{k,:} = A_{k,:} + B_{k,:}$, then* $\det(C) = \det A + \det B$.

16. Use the fact that if C is a square matrix with two identical columns, then $\det C = 0$ to prove that *if C is a square matrix with two identical rows, then* $\det C = 0$.

17. Prove that *the determinant of a square upper triangular matrix is the product of the entries on its diagonal* by expanding along the first column.

Answers

what is my matrix (swap)? The original matrix can be recovered by swapping the first two rows back:

$$\begin{bmatrix} 13 & -11 & -19 \\ 23 & 12 & 22 \\ -16 & 19 & -5 \end{bmatrix}$$

what is my matrix (scale)? The original matrix can be recoverd by scaling the third row by $\frac{1}{2}$ (the multiplicative inverse of 2):

$$\begin{bmatrix} 2 & -13 & -6 \\ -7 & -9 & 21 \\ -7 & -9 & 12 \end{bmatrix}$$

what is my matrix (replace)? The original matrix can be recoverd by replacing the first row by the first row plus negative three (the additive inverse of 3) times the third:

$$\begin{bmatrix} 2 & -15 & -4 \\ 3 & 5 & -8 \\ 3 & 8 & 1 \end{bmatrix}$$

determinant of a scale matrix A scale matrix is lower triangular with ones on the diagonal everywhere except the row that it scales, where the entry equals the scale factor. Since its determinant is the product of the entries on its diagonal, the determinant equals the scale factor.

determinant of a replace matrix A replace matrix is either lower triangular or upper triangular with ones on the main diagonal. Therefore its determinant is one. NOTE: This argument uses the fact in exercise 19 of section 3.4.

determinant of a scaled matrix $\det B = \det B^T = \det A^T = c \det A$ since B^T is the result of scaling a column of A^T and $\det M^T = \det M$.

3.6 Characterization of Square Matrices

For square matrices, all fourteen of the statements in theorems 5 and 6 are equivalent. A square matrix with a pivot in every row has a pivot in every column, and vice versa—end of justification. Square matrices have an additional property to discuss, though—invertibility. It turns out that, for a square matrix, the conditions in theorems 5 and 6 plus a couple that don't appear in those theorems are equivalent to invertibility. Consider the following.

1. M has a pivot position in every row and column.
2. $\det M \neq 0$.
3. M can be row reduced to I.
4. M is invertible.

You may or may not have considered these statements equivalent up to this point, and there is no harm done either way. It turns out they are equivalent to one another and equivalent to the statements in theorems 5 and 6. All this will be summarized in one last matrix characterization theorem, justified by the following narrative that shows (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (thm 5) and (thm 6) $\Rightarrow$ (1). Until the statement of the theorem, where this information will be repeated, assume that M is an $n \times n$ matrix.

Suppose M has a pivot position in every row and every column. **Record the elementary row operations, and more importantly the corresponding elementary matrices,** $E_1, E_2, \dots, E_k$**, that reduce** M **to any row echelon form** R**. Then** $E_k \cdots E_2 E_1 M = R$ **where** R **is in row echelon form. Because all elementary matrices are invertible,** $M = E_1^{-1} E_2^{-1} \cdots E_k^{-1} R$ **and therefore** $\det M = \det(E_1^{-1} E_2^{-1} \cdots E_k^{-1} R) = \det E_1^{-1} \cdot \det E_2^{-1} \cdots \det E_k^{-1} \cdot \det R$ **(a result of section 3.5**). Because the inverse of an elementary matrix is an elementary matrix itself and all elementary matrices have nonzero determinant, all the $\det E_i^{-1}$ are nonzero. Because M has a pivot position in every row, R must be upper triangular with nonzero entries (the pivots) on the diagonal, making $\det R$ equal to the product of these nonzero entries. Hence $\det R \neq 0$ and it follows that $\det M \neq 0$.

Suppose $\det M \neq 0$. The reduced row echelon form, R, can be represented by $R = E_k \cdots E_2 E_1 M$ for some elementary matrices $E_1, E_2, \dots, E_k$. Because $\det R = \det(E_k \cdots E_2 E_1 M) = \det E_k \cdots \det E_2 \cdot \det E_1 \cdot \det M$ and $\det M \neq 0$, $\det R$ must also have nonzero determinant. But the only reduced row echelon form of a square matrix with nonzero determinant is the identity (all others have a row of zeros, putting a zero on the main diagonal). Therefore M can be reduced to I.

Supposing M can be reduced to I, we have $E_k \cdots E_2 E_1 M = I$ for some elementary matrices $E_1, E_2, \dots, E_k$. Letting $E = E_k \cdots E_2 E_1$, we have $EM = I$. But elementary matrices are invertible, so E is invertible and therefore $M = E^{-1}$. Since E^{-1} is invertible (with inverse E), M is invertible.

Supposing M is invertible, let $L = R = M^{-1}$, proving the existence of matrices L and R such that $LM = I = MR$. By theorems 5 and 6, M has a pivot position in every row and column.

Supposing there is a matrix R such that $MR = I$, $\mathbf{v} = R\mathbf{b}$ is a solution of $M\mathbf{v} = \mathbf{b}$ since $M(R\mathbf{b}) = (MR)\mathbf{b} = I\mathbf{b} = \mathbf{b}$. Hence $M\mathbf{v} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$, and by theorem 6 M has a pivot position in each row. Since M is square, M has a pivot position in each column as well.

Crumpet 21: Proving Real Numbers are Equal

Every so often, it is convenient to prove that two real numbers, x and y, are equal by showing both $x \leq y$ and $x \geq y$. The only way x can be less than or equal to y and simultaneously greater than or equal to y is for x to equal y. This technique is implicitly used to justify part (ix) of theorem 7. Theorem 5 implies $M\mathbf{v} = \mathbf{b}$ has at most one solution (the number of solutions is less than or equal to one) and theorem 6 implies $M\mathbf{v} = \mathbf{b}$ has at least one solution (the number of solutions is greater than or equal to one). Together, then $M\mathbf{v} = \mathbf{b}$ has exactly one solution.

We now have justification for the following theorem.

Theorem 7. [Invertible Matrix Theorem] *Suppose M is an $n \times n$ matrix, and* $\mathbf{b}$ *and* $\mathbf{v}$ *have n entries. Then the following are equivalent.*

(i) *The columns of M are linearly independent.*

(ii) *The rows of M are linearly independent.*

(iii) *No column of M is a linear combination of the others.*

(iv) *No row of M is a linear combination of the others.*

(v) $M\mathbf{v} = \mathbf{0}$ *has only the trivial solution.*

(vi) *M has a pivot position in every column.*

(vii) *M has a pivot position in every row.*

(viii) $M\mathbf{v} = \mathbf{b}$ *has no free variables.*

(ix) $M\mathbf{v} = \mathbf{b}$ *has exactly one solution for every* $\mathbf{b}$.

(x) *M can be row reduced to I.*

(xi) *There is a matrix L such that $LM = I$.*

(xii) *There is a matrix R such that $MR = I$.*

(xiii) $\det M \neq 0$.

(xiv) *M is invertible.*

This theorem gives 13 ways to detect whether a square matrix is invertible, impressive in itself. But we can also draw two separate, significant conclusions from all this. Parts (xi) and (xii) suggest we only need to check that $AB = I$ ***or*** $BA = I$, ***not both*** as required by the definition, to conclude that B is the inverse of A. The theorem gives the other equality. Additionally, the bolded section of the justification, near the middle of page 106, provides an algorithm for calculating the determinant of a square matrix! Can you follow the instructions to compute the determinant of

$$\begin{bmatrix} 6 & 3 & 6 \\ -2 & 1 & -1 \\ 3 & 4 & 6 \end{bmatrix}?$$

Answer on page 109.

If you concluded in exercise 11 of section 1.5 that one row of a matrix could only be written as a linear combination of the others when the determinant of the matrix was zero, you were correct, and we finally have the theory to support it.

Key Concepts

characterization of invertible matrices see theorem 7.

algorithm for computing the determinant reduce the matrix to row echelon form, noting the row operations used. The product of the determinants of the inverses of the associated elementary matrices with the determinant of the reduced matrix is the desired determinant.

Exercises

1. The row operations that reduce a matrix A to

$$\begin{bmatrix} -5 & 15 & -10 \\ 0 & 12 & -14 \\ 0 & 0 & -2 \end{bmatrix}$$

are given. Find $\det A$. [$\mathbb{S}$]-300

(a) Ten row replacements.

(b) Five row replacements and three row swaps.

(c) Nine row replacements, a row scale by 6, and a row scale by 5.

(d) Four row replacements, a row scale by 10, and two row swaps.

2. The row operations that reduce a matrix A to

$$\begin{bmatrix} -2 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 13 & 10 & 5 & 0 \\ 0 & 2 & 14 & 6 \end{bmatrix}$$

are given. Find the possible values of $\det A$.

   (a) Row replacements only.

   (b) Row replacements and row swaps only. [𝕊]-301

   (c) Row replacements and row scales by 3, 10, and 14.

   (d) Row replacements, row swaps, and row scales.

3. The row operations that reduce a matrix A to

$$\begin{bmatrix} -21 & -2 & 6 \\ 0 & 0 & -5 \\ 0 & 0 & -9 \end{bmatrix}$$

are given. Find $\det A$.

   (a) Five row replacements and three row swaps.

   (b) Four row replacements, two row swaps, and a row scaling by -5.

   (c) 36 row replacements, 13 row swaps, and scaling by 12, -13, and $-\frac{17}{93}$.

4. Row reduce to a triangular matrix to compute the determinant.

   (a) $M = \begin{bmatrix} -3 & 10 \\ -15 & -10 \end{bmatrix}$

   (b) $M = \begin{bmatrix} -12 & 12 \\ 14 & 6 \end{bmatrix}$ [𝕊]-301

   (c) $M = \begin{bmatrix} 9 & -7 \\ 4 & 7 \end{bmatrix}$

   (d) $M = \begin{bmatrix} 16 & -3 & -2 \\ -8 & 4 & -2 \\ -8 & 1 & 2 \end{bmatrix}$

   (e) $M = \begin{bmatrix} -10 & 12 & -50 \\ 20 & -18 & 80 \\ -30 & 18 & -80 \end{bmatrix}$

   (f) $M = \begin{bmatrix} -11 & -15 & 4 \\ 8 & 9 & -4 \\ -3 & -3 & 2 \end{bmatrix}$ [𝕊]-301

   (g) $M = \begin{bmatrix} 3 & 90 & -308 & -6 \\ -3 & -140 & 484 & 10 \\ 6 & 210 & -737 & -16 \\ 3 & 70 & -231 & -4 \end{bmatrix}$ [𝕊]-301

   (h) $M = \begin{bmatrix} -80 & -161 & -18 & 55 \\ 80 & 154 & 27 & -66 \\ 0 & 0 & 9 & -11 \\ -24 & -49 & 27 & -22 \end{bmatrix}$

5. Compute the determinant using a judicious combination of row expansion, column expansion, and row reduction.

   (a) $\begin{bmatrix} -21 & 9 & 0 \\ -24 & 9 & 2 \\ -11 & 4 & 1 \end{bmatrix}$

   (b) $\begin{bmatrix} -1 & 5 & -1 \\ 0 & 6 & -1 \\ 4 & -10 & 2 \end{bmatrix}$ [𝕊]-302

   (c) $\begin{bmatrix} -5 & 2 & 1 \\ -9 & 1 & -1 \\ -32 & 4 & -3 \end{bmatrix}$

   (d) $\begin{bmatrix} -2 & -4 & -3 & 3 \\ -1 & -3 & 1 & 9 \\ 13 & 40 & -5 & -113 \\ 0 & 2 & 14 & 6 \end{bmatrix}$

   (e) $\begin{bmatrix} -9 & 12 & -87 & -25 \\ -5 & 5 & -47 & -14 \\ -3 & 8 & -28 & 12 \\ 1 & -1 & 10 & 3 \end{bmatrix}$

6. Is the matrix invertible? Explain.

   (a) $\begin{bmatrix} -6 & -3 \\ 0 & 19 \end{bmatrix}$

   (b) $\begin{bmatrix} 3 & -7 \\ 0 & -20 \end{bmatrix}$

   (c) $\begin{bmatrix} 1 & 1 \\ -5 & -4 \end{bmatrix}$

   (d) $\begin{bmatrix} 1 & 1 & 2 \\ 0 & 3 & 11 \\ 0 & 3 & -14 \end{bmatrix}$

   (e) $\begin{bmatrix} 1 & 1 & 3 \\ 4 & 7 & 11 \\ 0 & 3 & 20 \end{bmatrix}$ [𝕊]-302

   (f) $\begin{bmatrix} 3 & 1 & -2 \\ -7 & -16 & -9 \\ 8 & 5 & -3 \end{bmatrix}$

7. Suppose M is not invertible yet there is a matrix R such that $MR = I$. How is this possible?

8. Suppose M is square and $3M_{:,2} = 2M_{:,1} - 8M_{:,5} + \frac{1}{2}M_{:,6}$. What is $\det M$?

9. Suppose the rows of M are linearly independent but M is not invertible. How can this be?

10. Explain why a matrix with a pivot position in every row and every column must be invertible.

11. Suppose G is square and $G\mathbf{v} = \mathbf{b}$ is inconsistent for some vector $\mathbf{b}$. What can you say about solutions of $G\mathbf{v} = \mathbf{0}$? [𝔸]-351

12. If G is square and $G\mathbf{v} = \mathbf{0}$ has infinitely many solutions, what can you say about solutions of $G\mathbf{v} = \mathbf{b}$?

13. If M is invertible, then the rows of M^T are linearly indepedent. Explain why. [𝔸]-351

14. If H is 7×7 and $H\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b}$, how many pivot positions does H have?

15. If a square matrix B cannot be reduced to the identity matrix, what can you say about [𝔸]-351

    (a) its columns?

    (b) the equation $B\mathbf{v} = \mathbf{0}$?

(c) the equation $AB = I$?

16. Describe the row echelon form of an invertible matrix.

17. When the determinant of an $n \times n$ matrix is zero, (select all that apply) [𝔸]-351

(a) exactly one row is a linear combination of the others.

(b) every row is a linear combination of the others.

(c) each row after the first one is a linear combination of the rows above it.

(d) any linear combination of the n rows sums to zero.

(e) at least one row is a linear combination of the others.

(f) its inverse is the zero matrix.

(g) it has no inverse.

18. Recall that $\lambda, \mathbf{v}$ is an eigenpair for M whenever $\mathbf{v} \neq \mathbf{0}$ yet $(M - \lambda I)\mathbf{v} = \mathbf{0}$. Use theorem 7 to prove that the following statements are equivalent.

(a) λ is an eigenvalue of M.

(b) The rows of $M - \lambda I$ are linearly dependent.

(c) $\det(M - \lambda I) = 0$.

Answers

determinant The instructions are, in brief: **Record the elementary row operations, and more importantly [note] the corresponding elementary matrices, $E_1, E_2, \ldots, E_k$, that reduce M to any row echelon form. Then** $\det M = \det E_1^{-1} \cdot \det E_2^{-1} \cdots \det E_k^{-1} \cdot \det R$**.**

Recording the elementary row operations during row reduction:

$$\begin{bmatrix} 6 & 3 & 6 \\ -2 & 1 & -1 \\ 3 & 4 & 6 \end{bmatrix} \underset{M_{3,:} \to -2M_{3,:}}{\overset{M_{2,:} \to 3M_{2,:}}{\longrightarrow}} \begin{bmatrix} 6 & 3 & 6 \\ -6 & 3 & -3 \\ -6 & -8 & -12 \end{bmatrix} \underset{M_{3,:} \to M_{3,:}+M_{1,:}}{\overset{M_{2,:} \to M_{2,:}+M_{1,:}}{\longrightarrow}}$$

$$\begin{bmatrix} 6 & 3 & 6 \\ 0 & 6 & 3 \\ 0 & -5 & -6 \end{bmatrix} \overset{M_{2,:} \to M_{2,:}+M_{3,:}}{\longrightarrow} \begin{bmatrix} 6 & 3 & 6 \\ 0 & 1 & -3 \\ 0 & -5 & -6 \end{bmatrix} \overset{M_{3,:} \to M_{3,:}+5M_{2,:}}{\longrightarrow} \begin{bmatrix} 6 & 3 & 6 \\ 0 & 1 & -3 \\ 0 & 0 & -21 \end{bmatrix}$$

The determinant of $\begin{bmatrix} 6 & 3 & 6 \\ -2 & 1 & -1 \\ 3 & 4 & 6 \end{bmatrix}$ is the determinant of $\begin{bmatrix} 6 & 3 & 6 \\ 0 & 1 & -3 \\ 0 & 0 & -21 \end{bmatrix}$, which is $6 \cdot 1 \cdot -21 = -126$, times the determinants of the inverse elementary matrices:

$$\begin{vmatrix} 6 & 3 & 6 \\ -2 & 1 & -1 \\ 3 & 4 & 6 \end{vmatrix} = (-126)(1)(1)(1)(1)\left(-\frac{1}{2}\right)\left(\frac{1}{3}\right)$$

$$= 21.$$

3.7 The Inverse Revisited

As if we haven't already extracted enough information from theorem 7, we also have the rather significant following theorem as a consequence.

Theorem 8. [Determinant of a Product] *If A and B are $n \times n$ matrices, then* $\det(AB) = \det A \cdot \det B$.

Proof. First suppose AB is noninvertible. By theorem 106, $\det(AB) = 0$. If both A and B are invertible, then $(AB)(B^{-1}A^{-1}) = I$, so AB is invertible. Therefore we must have that either A or B is noninvertible, from which it follows $\det A = 0$ or $\det B = 0$. Either way, $\det A \cdot \det B = 0$ and we have shown $\det(AB) = \det A \cdot \det B$. Now suppose AB is invertible, and let $M = (AB)^{-1}$. Then $I = (AB)M = A(BM)$, so $A^{-1} = BM$ and A is invertible. As in the justification of 3.⇒4. on page 106, we may therefore write A as a product of elementary matrices, $E_k^{-1} \cdots E_2^{-1} E_1^{-1}$. Hence $\det(AB) = \det(E_k^{-1} \cdots E_2^{-1} E_1^{-1} B) = (\det E_k^{-1} \cdots \det E_2^{-1} \det E_1^{-1}) \det B = \det A \det B$. □

As a direct consequence, we can relate the determinants of inverse matrices. If M is invertible, then $\det M \cdot \det M^{-1} = \det I = 1$ and therefore $\det M^{-1} = \frac{1}{\det M}$.

There is more! The proof of theorem 7 also provides an algorithm for finding the inverse of a matrix. Given that M is invertible, it is reducible to the identity matrix, meaning there are elementary matrices $E_1, E_2, \ldots, E_k$ such that $E_1, E_2, \ldots, E_k M = I$. Therefore $M^{-1} = E_1 E_2 \cdots E_k I$, so the same sequence of elementary row operations that reduces M to the identity also transforms I into M^{-1}! Hence, if we augment M with the identity matrix and reduce to reduced row echelon form, the augmented columns will hold M^{-1}. To illustrate, let

$$M = \begin{bmatrix} 3 & 0 & 5 & 0 \\ 5 & 1 & 0 & 2 \\ 6 & 2 & 0 & 7 \\ 0 & 0 & -1 & -2 \end{bmatrix}$$

Augmenting the identity and reducing,

$$\begin{bmatrix} 3 & 0 & 5 & 0 & 1 & 0 & 0 & 0 \\ 5 & 1 & 0 & 2 & 0 & 1 & 0 & 0 \\ 6 & 2 & 0 & 7 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & -2 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{M_{1,:} \leftrightarrow M_{3,:}} \begin{bmatrix} 6 & 2 & 0 & 7 & 0 & 0 & 1 & 0 \\ 5 & 1 & 0 & 2 & 0 & 1 & 0 & 0 \\ 3 & 0 & 5 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -2 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{M_{1,:} \to M_{1,:} - M_{2,:}}$$

$$\begin{bmatrix} 1 & 1 & 0 & 5 & 0 & -1 & 1 & 0 \\ 5 & 1 & 0 & 2 & 0 & 1 & 0 & 0 \\ 3 & 0 & 5 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -2 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow[M_{3,:} \to M_{3,:} - 3M_{1,:}]{M_{2,:} \to M_{2,:} - 5M_{1,:}} \begin{bmatrix} 1 & 1 & 0 & 5 & 0 & -1 & 1 & 0 \\ 0 & -4 & 0 & -23 & 0 & 6 & -5 & 0 \\ 0 & -3 & 5 & -15 & 1 & 3 & -3 & 0 \\ 0 & 0 & -1 & -2 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{M_{2,:} \to M_{2,:} - M_{3,:}}$$

$$\begin{bmatrix} 1 & 1 & 0 & 5 & 0 & -1 & 1 & 0 \\ 0 & -1 & -5 & -8 & -1 & 3 & -2 & 0 \\ 0 & -3 & 5 & -15 & 1 & 3 & -3 & 0 \\ 0 & 0 & -1 & -2 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{M_{2,:} \to -1M_{2,:}} \begin{bmatrix} 1 & 1 & 0 & 5 & 0 & -1 & 1 & 0 \\ 0 & 1 & 5 & 8 & 1 & -3 & 2 & 0 \\ 0 & -3 & 5 & -15 & 1 & 3 & -3 & 0 \\ 0 & 0 & -1 & -2 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{M_{3,:} \to M_{3,:} + 3M_{2,:}}$$

$$\begin{bmatrix} 1 & 1 & 0 & 5 & 0 & -1 & 1 & 0 \\ 0 & 1 & 5 & 8 & 1 & -3 & 2 & 0 \\ 0 & 0 & 20 & 9 & 4 & -6 & 3 & 0 \\ 0 & 0 & -1 & -2 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{M_{3,:} \to M_{3,:} + 20M_{4,:}} \begin{bmatrix} 1 & 1 & 0 & 5 & 0 & -1 & 1 & 0 \\ 0 & 1 & 5 & 8 & 1 & -3 & 2 & 0 \\ 0 & 0 & 0 & -31 & 4 & -6 & 3 & 20 \\ 0 & 0 & -1 & -2 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{M_{3,:} \leftrightarrow M_{4,:}}$$

$$\left[\begin{array}{cccccccc} 1 & 1 & 0 & 5 & 0 & -1 & 1 & 0 \\ 0 & 1 & 5 & 8 & 1 & -3 & 2 & 0 \\ 0 & 0 & -1 & -2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -31 & 4 & -6 & 3 & 20 \end{array}\right] \underset{M_{4,:}\to\frac{-1}{31}M_{4,:}}{\overset{M_{3,:}\to -1M_{3,:}}{\longrightarrow}} \left[\begin{array}{cccccccc} 1 & 1 & 0 & 5 & 0 & -1 & 1 & 0 \\ 0 & 1 & 5 & 8 & 1 & -3 & 2 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & -\frac{4}{31} & \frac{6}{31} & -\frac{3}{31} & -\frac{20}{31} \end{array}\right] \underset{M_{1,:}\to M_{1,:}-M_{2,:}}{\overset{M_{2,:}\to M_{2,:}-5M_{3,:}}{\longrightarrow}}$$

$$\left[\begin{array}{cccccccc} 1 & 0 & 0 & 7 & -1 & 2 & -1 & -5 \\ 0 & 1 & 0 & -2 & 1 & -3 & 2 & 5 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & -\frac{4}{31} & \frac{6}{31} & -\frac{3}{31} & -\frac{20}{31} \end{array}\right] \overset{M_{3,:}\to M_{3,:}-2M_{4,:}}{\longrightarrow} \left[\begin{array}{cccccccc} 1 & 0 & 0 & 7 & -1 & 2 & -1 & -5 \\ 0 & 1 & 0 & -2 & 1 & -3 & 2 & 5 \\ 0 & 0 & 1 & 0 & \frac{8}{31} & -\frac{12}{31} & \frac{6}{31} & \frac{9}{31} \\ 0 & 0 & 0 & 1 & -\frac{4}{31} & \frac{6}{31} & -\frac{3}{31} & -\frac{20}{31} \end{array}\right]$$

$$\underset{M_{1,:}\to M_{1,:}-7M_{4,:}}{\overset{M_{2,:}\to M_{2,:}+2M_{4,:}}{\longrightarrow}} \left[\begin{array}{cccccccc} 1 & 0 & 0 & 0 & -\frac{3}{31} & \frac{20}{31} & -\frac{10}{31} & -\frac{15}{31} \\ 0 & 1 & 0 & 0 & \frac{23}{31} & -\frac{81}{31} & \frac{56}{31} & \frac{115}{31} \\ 0 & 0 & 1 & 0 & \frac{8}{31} & -\frac{12}{31} & \frac{6}{31} & \frac{9}{31} \\ 0 & 0 & 0 & 1 & -\frac{4}{31} & \frac{6}{31} & -\frac{3}{31} & -\frac{20}{31} \end{array}\right]$$

so

$$M^{-1} = \frac{1}{31}\left[\begin{array}{cccc} -3 & 20 & -10 & -15 \\ 23 & -81 & 56 & 115 \\ 8 & -12 & 6 & 9 \\ -4 & 6 & -3 & -20 \end{array}\right]$$

Crumpet 22: Inverses via Row Reduction

We could have seen that inverses could be computed with the help of row reduction long ago. After all, if A is an $n \times n$ matrix and B is its inverse, then $AB = I$. By thinking of this product one column at a time, this means

$$AB_{:,1} = I_{:,1},\ AB_{:,2} = I_{:,2},\ \ldots,\ AB_{:,n} = I_{:,n}.$$

Solving these equations for the $B_{:,i}$ could be done one at a time by row reduction. Putting the solutions together into a matrix would give B. Reducing A n times would be repetitive and time consuming, though. Better, the solutions could be found simultaneously by augmenting all of the $I_{:,i}$ together—in effect, augmenting the identity matrix—and reducing once (the algorithm presented in this section).

While this process is still tedious for large matrices, it certainly beats the alternative of using formula (1.6.2). Ironically the ideas presented recently give us the tools to finally prove that (1.6.2) correctly computes the inverse. Let M be an $n \times n$ matrix and consider modifying M by replacing row j with a copy of row i, $i \neq j$. Call the modified matrix $\tilde{M}$. Then

$$\left|\tilde{M}\right| = (-1)^{j+1}\tilde{M}_{j,1}\left|\tilde{M}_{\setminus j,1}\right| + (-1)^{j+2}\tilde{M}_{j,2}\left|\tilde{M}_{\setminus j,2}\right| + \cdots + (-1)^{j+n}\tilde{M}_{j,n}\left|\tilde{M}_{\setminus j,n}\right|.$$

But $\tilde{M}_{j,k} = M_{i,k}$ and $\tilde{M}_{\setminus j,k} = M_{\setminus j,k}$ by construction, so

$$\left|\tilde{M}\right| = (-1)^{j+1}M_{i,1}\left|M_{\setminus j,1}\right| + (-1)^{j+2}M_{i,2}\left|M_{\setminus j,2}\right| + \cdots + (-1)^{j+n}M_{i,n}\left|M_{\setminus j,n}\right|.$$

On the other hand, $\left|\tilde{M}\right| = 0$ since $\tilde{M}$ has two identical rows. We conclude that

$$(-1)^{j+1}M_{i,1}\left|M_{\setminus j,1}\right| + (-1)^{j+2}M_{i,2}\left|M_{\setminus j,2}\right| + \cdots + (-1)^{j+n}M_{i,n}\left|M_{\setminus j,n}\right| \tag{3.7.1}$$

equals 0 whenever $i \neq j$. Observe that when $i = j$, (3.7.1) is $\det M$ expanded along row i (or row j depending on your perspective). The proof of formula (1.6.2) then lies in noticing that for any square matrix A, the entries of the product

$$A \cdot \mathrm{adj}A$$

all take the form (3.7.1). Accordingly $A \cdot \mathrm{adj}A = (\det A)I$ for any square matrix A. If A is invertible, $\det A \neq 0$ and we have $A \cdot \frac{1}{\det A}\mathrm{adj}A = I$, so $A^{-1} = \frac{1}{\det A}\mathrm{adj}A$.

Another place where row reduction could help ease an earlier burden is finding eigenvectors. Unless you happened to work through exercise 5 of section 2.3, the last time you were asked to compute an eigenvector, you were expected to write out a linear system of equations without using matrix notation and to solve the system using elimination or substitution, not row operations. With the introduction of the parametric vector form for writing solution sets of linear systems with infinitely many solutions, there is no reason not to apply matrix techniques to the task of finding eigenvectors. Can you use row reduction to find the eigenvectors of

$$M = \begin{bmatrix} -17 & 49 \\ -21 & 53 \end{bmatrix}$$

given that its eigenvalues are 4 and 32? Answer on page 113.

Key Concepts

determinant of an inverse if M is invertible, $\det M^{-1} = \frac{1}{\det M}$.

determinant of a product if A and B are $n \times n$ matrices, then $\det(AB) = \det A \cdot \det B$.

inverses by row reduction if A is invertible, then $\begin{bmatrix} A & I \end{bmatrix}$ row reduces to $\begin{bmatrix} I & A^{-1} \end{bmatrix}$.

eigenvectors by row reduction if λ is an eigenvalue of M, then corresponding eigenvectors can be found by row reducing $M - \lambda I$.

Exercises

1. Find the inverse by row reduction.

   (a) $\begin{bmatrix} -3 & 10 \\ -15 & -10 \end{bmatrix}$

   (b) $\begin{bmatrix} -12 & 12 \\ 14 & 6 \end{bmatrix}$

   (c) $\begin{bmatrix} 9 & -7 \\ 4 & 7 \end{bmatrix}$ [𝕊]-302

   (d) $\begin{bmatrix} 16 & -3 & -2 \\ -8 & 4 & -2 \\ -8 & 1 & 2 \end{bmatrix}$ [𝕊]-302

   (e) $\begin{bmatrix} -10 & 12 & -50 \\ 20 & -18 & 80 \\ -30 & 18 & -80 \end{bmatrix}$

   (f) $\begin{bmatrix} -11 & -15 & 4 \\ 8 & 9 & -4 \\ -3 & -3 & 2 \end{bmatrix}$

   (g) $\begin{bmatrix} 3 & 90 & -308 & -6 \\ -3 & -140 & 484 & 10 \\ 6 & 210 & -737 & -16 \\ 3 & 70 & -231 & -4 \end{bmatrix}$

2. If $\det M = 2$ and $\det R = \frac{1}{3}$ and M and R are the same size, find [𝕊]-303

   (a) $\det(MR^T)$

   (b) $\det(M^{-1}R)$

   (c) $\det(MR^{-1})^T$

3. Suppose L, A, M, B are square matrices such that $\det(LA) = 6$, $\det(AM) = 24$, and $\det(MB) = 48$. Find

   (a) $\det(LA^T)$

   (b) $\det(LM^{-1})$

   (c) $\det(LAMB)$

   (d) $\det(LB)$

4. For a square matrix M, explain why the determinant of $M^T M$ must be nonnegative.

5. Suppose M is invertible. Explain why PMP^{-1} is invertible for any (invertible) matrix P.

6. Support the claim that *the product of invertible matrices is invertible*.

7. Explain why $\det(PMP^{-1}) = \det M$ for any matrices M and P, assuming both sides of the equation are defined.

8. Suppose [𝔸]-351

   $$AA^T = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}.$$

   (a) Is A necessarily invertible?

   (b) If A is square, is A necessarily invertible?

9. If λ is an eigenvalue of M, what can you say about the pivot positions of $M - \lambda I$?

10. Suppose $M - cI$ has linearly independent columns. Can c be an eigenvalue of M? Explain. [𝔸]-351

11. Use row reduction to find the eigenvectors corresponding to the given eigenvalue. Write your answer in parametric vector form.

    (a) $A = \begin{bmatrix} 3 & -10 \\ 8 & -15 \end{bmatrix}; \lambda = -5$

    (b) $A = \begin{bmatrix} -4 & 2 \\ -16 & 8 \end{bmatrix}; \lambda = 0$

    (c) $A = \begin{bmatrix} 2 & 4 \\ -3 & -4 \end{bmatrix}; \lambda = -1 - i\sqrt{3}$ [𝔸]-351

    (d) $A = \begin{bmatrix} 9 & 1 & -5 \\ 33 & 17 & -25 \\ 36 & 12 & -24 \end{bmatrix}; \lambda = 6$

(e) $A = \begin{bmatrix} 14 & 9 & 18 \\ 12 & 17 & -18 \\ 12 & -9 & 8 \end{bmatrix}$; $\lambda = 26$ [A]-352

(f) $A = \begin{bmatrix} -5 & 6 & -12 \\ 7 & -8 & 16 \\ 5 & -6 & 12 \end{bmatrix}$; $\lambda = -2$

(g) $A = \begin{bmatrix} 45 & -51 & -24 & -60 \\ 15 & 107 & 18 & 0 \\ 15 & 17 & 98 & 20 \\ -30 & -34 & -16 & 50 \end{bmatrix}$; $\lambda = 90$ [A]-352

(h) $A = \begin{bmatrix} -10 & -2 & -2 & -6 \\ -26 & -79 & -1 & -93 \\ -116 & -106 & -4 & -138 \\ 26 & 61 & 1 & 75 \end{bmatrix}$; $\lambda = -18$

Answers

eigenvectors if λ is an eigenvalue of M, then the associated eigenvector, $\mathbf{v}$, satisfies $(M - \lambda I)\mathbf{v} = \mathbf{0}$. For

$$M = \begin{bmatrix} -17 & 49 \\ -21 & 53 \end{bmatrix}$$

and $\lambda = 4$, that means the unknown eigenvector satisfies

$$\begin{bmatrix} -21 & 49 \\ -21 & 49 \end{bmatrix} \mathbf{v} = \mathbf{0}.$$

This system can be solved by reducing the augmented matrix

$$\begin{bmatrix} -21 & 49 & 0 \\ -21 & 49 & 0 \end{bmatrix}.$$

Subtracting row 1 from row 2 yields

$$\begin{bmatrix} -21 & 49 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

v_2 is a free variable and $v_1 = \frac{-49}{-21} v_2 = \frac{7}{3} v_2$. In parametric vector form,

$$\mathbf{v} = r \begin{bmatrix} \frac{7}{3} \\ 1 \end{bmatrix}$$

or equivalently

$$\mathbf{v} = r \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

for any r. Speeding up the process for the eigenvalue $\lambda = 32$, we need to reduce the augmented matrix

$$\begin{bmatrix} -49 & 49 & 0 \\ -21 & 21 & 0 \end{bmatrix}.$$

Again the second row disappears with one row operation, leaving

$$\begin{bmatrix} -49 & 49 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

from which we deduce $v_1 = v_2$. The solution is therefore

$$\mathbf{v} = r \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

for any r.

Part II

Matrix Abstraction

Chapter 4

Vector Spaces and Inner Product Spaces

Abstraction is at the heart of most of mathematics. It is an essential vehicle for the development of new ideas. Take natural numbers (one, two, three, and so on), for example. These numbers have a "natural" meaning—quantity. If you have a number of objects before you, then there are one or two or maybe twenty-two. The objects are there. They can be counted. But what does it mean to "count" the number of objects before you when there are none? The very idea is an abstraction of the notion of counting, and it leads to the number zero. To the natural numbers, we add this number we call zero, and say that it represents the number of objects you have when you have none. You cannot explain zero in the same concrete way you can explain one or two or three. It requires an intellectual leap in one's understanding of counting.

The use of variables requires another intellectual leap. It is one thing to say that $4+5 = 5+4$, but it is quite another to say that $x + y = y + x$. The unspecified quantities are abstractions of numbers. Much like zero, they represent something that is not readily available to see or put in other concrete terms. The very idea of using unspecified quantities gives rise to an entire branch of mathematics—algebra! Many branches of mathematics revolve around similar abstractions. A body of objects or ideas is stripped down to its essence, which then gives rise to similar but new objects or ideas.

Thinking of the numbers one, two, three, and so on as counting numbers allows the abstraction of "counting" zero objects. Using symbols other than numbers to stand for numbers allows the abstraction of unspecified quantities in an expression. Abstraction provides a new perspective from which new mathematics can bloom. This idea is the foundation for many branches of mathematics. Abstract algebra is built upon abstraction of binary operators such as addition and multiplication, both of which are associative, admit an identity element and inverses, and are closed on certain sets of real numbers. Topology is built upon abstraction of open intervals of the real line, arbitrary unions of which are also open, and closed intervals of the real line, finite intersections of which are also closed. Non-Euclidean geometry is built upon abstraction of the idea of a line. Analysis is built upon abstraction of finite quantity and size to infinite quantity and infinitesimal size. At some level, each branch of mathematics is based upon abstraction.

This part of the book proceeds in this vein. The essential ingredients of objects and ideas already explored and understood in a concrete way (vectors, matrix multiplication, and dot product to be specific) will be extracted from their concrete settings, opening doors to new mathematics.

4.1 Vector Spaces and Span

Back in section 1.5, the definition of linear combination contained a proviso: "let S be a set of objects on which addition and scalar mutliplication are defined." At the time several examples of such sets were given, but not much was made of other properties the set should have. Then in section 3.3, the definition of linear independence contained an extended proviso: "[l]et S be a set of objects on which addition and scalar mutliplication are defined and which contains an additive identity, called 0." Existence of an additive identity was necessary to write the equation in the definition. Yet again, little was made of additional properties the set S should possess.

Implicit in the assumptions that addition and scalar multiplication are defined is that sums and scalar products of objects in the set S are also in the set S. This property is called closure of the operation. The set $T = \{t, t^2, 1 + t^2\}$

is not closed under addition nor scalar multiplication. Can you verify this? Answer on page 123. The set $S = \{\alpha t + \beta t^2 + \gamma(1 + t^2) : \alpha, \beta, \gamma \in \mathbb{R}\}$ containing all linear combinations of elements of T (with scalars from the set of real numbers) is closed under addition and scalar multiplication. Can you verify this? Answer on page 123. Closure is an essential (and until now unmentioned) property of linear combinations and linear independence, and there are others.

We are so used to the basic properties of real numbers, such as associativity and commutativity, it is easy to take the subtleties of computations for granted. In order for $r(\alpha t + \beta t^2 + \gamma(1 + t^2))$ to equal $(r\alpha)t + (r\beta)t^2 + (r\gamma)(1 + t^2)$, for example, we need to distribute the r and associate scalars. Distribution alone yields $r(\alpha t) + r(\beta t) + r(\gamma(1 + t^2))$. The distributive and associative properties are critical to the claim that S is closed under addition and scalar multiplication. In order for $\left[\alpha_1 t + \beta_1 t^2 + \gamma_1(1 + t^2)\right] + \left[\alpha_2 t + \beta_2 t^2 + \gamma_2(1 + t^2)\right]$ to equal $(\alpha_1 + \alpha_2)t + (\beta_1 + \beta_2)t^2 + (\gamma_1 + \gamma_2)(1 + t^2)$ we need a slightly different distributive property, an associative property, and commutativity of addition! These properties are also essential.

It is time to make explicit all the properties of matrices we have taken for granted when discussing linear combinations. This exercise will provide an abstract footing from which to explore, revealing similarities between certain sets of objects that could otherwise easily go unnoticed.

A nonempty set V on which addition and scalar multiplication are defined is called a **vector space** if for every $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V and all scalars s, t

1. $\mathbf{u} + \mathbf{v}$ is in V (V is closed under addition)
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (addition is commutative)
3. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ (addition is associative)
4. there is an element $\mathbf{0}$ in V such that $\mathbf{0} + \mathbf{u} = \mathbf{u}$ (an additive identity exists)
5. there is an element $-\mathbf{u}$ in V such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$. (every element has an additive inverse)
6. $s\mathbf{v}$ is in V (V is closed under scalar multiplication)
7. $1\mathbf{v} = \mathbf{v}$
8. $s(\mathbf{u} + \mathbf{v}) = s\mathbf{u} + s\mathbf{v}$ (scalars distribute over elements of V)
9. $(s + t)\mathbf{u} = s\mathbf{u} + t\mathbf{u}$ (elements of V distribute over scalars)
10. $s(t\mathbf{u}) = (st)\mathbf{u}$ (scalar multiplication is associative)

The set of all $n \times 1$ matrices (vectors of size n) with real entries is the model from which this list is derived. We use the symbol $\mathbb{R}$ for the set of all real numbers, and use the symbol $\mathbb{R}^n$ for the set of all ordered lists of n real numbers, their representation as $n \times 1$ matrices with real entries giving meaning to the operations of addition and scalar multiplication. For $V = \mathbb{R}^n$, properties 2,3,4,8,9, and 10 are taken directly from the theorems of section 3.1; property 5 is addressed in exercise 16 of the same section; and the other three follow from properties of real numbers. Property 7 is clear in $\mathbb{R}^n$ since $1 \cdot r = r$ for any real number r. Closure (properties 1 and 6) are also clear in $\mathbb{R}^n$ since sums and products of real numbers are real. Hence, these ten properties describe the essential features of $\mathbb{R}^n$.

The set $\mathbb{P}_n(\mathbb{R})$, the set of all polynomials with real coefficients and degree n or less, is a vector space[1]. We will verify this shortly by showing that this set satisfies the 10 properties of a vector space. Proof that other sets form vector spaces will be requested in the exercises. Elements of any vector space V are called **vectors** (even if they are matrices, sequences, or polynomials). In this abstract sense, even functions are vectors! Proof that $\mathbb{P}_n(\mathbb{R})$ is a vector space is mostly an exercise of *observing* that polynomials have the right properties rather than *demonstrating* that polynomials have the right properties.

1. The sum of two polynomials of degree n or less is a polynomial of degree n or less. (closure under addition)
2. For any polynomials $p(x)$ and $q(x)$, $p(x) + q(x) = q(x) + p(x)$. (addition is commutative)
3. For any polynomials $p(x)$, $q(x)$, and $r(x)$, $p(x) + (q(x) + r(x)) = (p(x) + q(x)) + r(x)$. (addition is associative)

[1] with the understanding that adding polynomials and multiplying polynomials by real numbers work as in high school algebra.

4. The polynomial $z(x) = 0$ is in $\mathbb{P}_n(\mathbb{R})$ and has the property that $z(x) + q(x) = q(x)$ for any polynomial $q(x)$. (an additive identity exists)
5. Given any polynomial $p(x)$, $p(x) + (-1 \cdot p(x)) = z(x)$ and $-1 \cdot p(x)$ is in $\mathbb{P}_n(\mathbb{R})$. (every element has an additive inverse)
6. For any polynomial $p(x)$ of degree n or less, $rp(x)$ is a polynomial of degree n or less. (closure under scalar multiplication)
7. $1 \cdot p(x) = p(x)$ for any polynomial $p(x)$.
8. $r(p(x) + q(x)) = rp(x) + rq(x)$ for any real number r and polynomials $p(x)$ and $q(x)$. (real numbers distribute over polynomials)
9. $(r + s)p(x) = rp(x) + sp(x)$ for any real numbers r and s and polynomial p. (polynomials distribute over real numbers)
10. $r(sp(x)) = (rs)p(x)$ for any real numbers r and s and polynomial p. (scalar multiplication is associative)

If the properties of polynomials were not so well known, each of these points would need to be accompanied by reference to a theorem or axiom for support. Even with great familiarity, the same four properties that derived from properties of real numbers (1,4,5,6) require further attention. Closure (properties 1 and 6) cannot be taken for granted. These properties only follow because scalar multiplication and polynomial addition do not increase the degree of a polynomial. Properties 4 and 5 assert that an additive identity is in the set and for every element of the set, its additive inverse is also in the set. These properies are not automatic for arbitrary subsets of polynomials. It is critical to note that $\mathbb{P}_n(\mathbb{R})$ contains them.

Coincidentally, given a subset H of any vector space V, the same four properties are the only ones that need to be shown true (in H) to prove that H is itself a vector space. The other six properties are inherited by H through the fact that they hold for *all* elements of V (including those in H). In fact property 5 can be deduced once properties 1 and 6 have been established.

Suppose H is a subset of a vector space V, properties 1 and 6 hold for H, and $\mathbf{h}$ is a particular but arbitrary element of H. Because H is closed under scalar multiplication (property 6), $-1 \cdot \mathbf{h}$ is in H. Because H is closed under addition (property 1), $\mathbf{h} + (-1 \cdot \mathbf{h})$ is in H. But $\mathbf{h} + (-1 \cdot \mathbf{h}) = (1 + (-1))\mathbf{h} = 0\mathbf{h}$ by property 9. If it were true that $0\mathbf{h} = \mathbf{0}$ (like it is in $\mathbb{R}^n$), we would be done as $-1 \cdot \mathbf{h}$ would then have been shown to be an additive inverse of $\mathbf{h}$ in H. Can you prove that $0\mathbf{h} = \mathbf{0}$ using only the ten properties of a vector space? Answer on page 123. Hence a subset of a vector space is a vector space itself if it satisfies properties 1, 4, and 6 (closure and containment of the zero vector).

Any subset of a vector space that is itself a vector space is called a **subspace**. One common way to define a subspace is through the collection of all linear combinations of a set of vectors (reminiscent of how $T = \{t, t^2, 1 + t^2\}$ is not closed under addition or multiplication but the collection of all linear combinations of elements of T is). Such a set is called the **span** of those vectors. To be precise, if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are elements of a vector space (vectors), the span of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, denoted $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, is given by

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k : c_1, c_2, \ldots, c_k \text{ are scalars}\}.$$

Can you show that given any vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ of a vector space, $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a subspace? Answer on page 123. For consistency with the notion of exercise 26 of this section and the nonemptiness of vector spaces (and the concept of a basis, which will be studied later), the span of the empty set is $\{\mathbf{0}\}$. That is, $\text{span}\{\} = \{\mathbf{0}\}$.

Key Concepts

vector space A set V on which addition and scalar multiplication are defined is called a **vector space** if for every $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V and all scalars (real numbers or complex numbers) s, t

1. $\mathbf{u} + \mathbf{v}$ is in V (V is closed under addition)
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (addition is commutative)
3. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ (addition is associative)
4. there is an element $\mathbf{0}$ in V such that $\mathbf{0} + \mathbf{u} = \mathbf{u}$ (an additive identity exists)

5. there is an element $-\mathbf{u}$ in V such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$. (every element has an additive inverse)
6. $s\mathbf{v}$ is in V (V is closed under scalar multiplication)
7. $1\mathbf{v} = \mathbf{v}$
8. $s(\mathbf{u} + \mathbf{v}) = s\mathbf{u} + s\mathbf{v}$ (scalars distribute over elements of V)
9. $(s + t)\mathbf{u} = s\mathbf{u} + t\mathbf{u}$ (elements of V distribute over scalars)
10. $s(t\mathbf{u}) = (st)\mathbf{u}$ (scalar multiplication is associative)

vector an element of a vector space.

span for vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ of a vector space, the span of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, denoted span$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, is the collection of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$. That is,

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k : c_1, c_2, \ldots, c_k \text{ are scalars}\}.$$

Additionally, span$\{\} = \{\mathbf{0}\}$.

subspace a subset of a vector space that is itself a vector space.

subspace conditions a subset of a vector space is a subspace if it contains $\mathbf{0}$ and is closed under addition and scalar multiplication.

span as subspace given any subset H of a vector space V, span$H$ is a subspace of V.

Notation

Some subsets of the complex numbers are denoted as follows.
$\mathbb{Z} = \{z : z \text{ is an integer}\}$
$\mathbb{Z}^+ = \{z : z \text{ is a positive integer}\}$
$\mathbb{Q} = \{q : q \text{ is a rational number}\}$
$\mathbb{R} = \{r : r \text{ is a real number}\}$
$\mathbb{C} = \{c : c \text{ is a complex number}\}$
Some sets, each of which has a natural definition as a vector space, are denoted as follows. $\mathbb{F}$ is taken to be one of[2] $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$, and D is taken to be a subset of $\mathbb{R}$.
$\mathbb{R}^n = \{\mathbf{v} : \mathbf{v} \text{ is an ordered list of } n \text{ real numbers}\}$
$\mathbb{R}^\mathbb{N} = \{s : s \text{ is a sequence of real numbers}\}$
$\mathcal{M}_{m\times n}(\mathbb{F}) = \{M : M \text{ is an } m \times n \text{ matrix with entries in the set } \mathbb{F}\}$
$\mathbb{P}_n(\mathbb{F}) = \{p : p \text{ is a polynomial of degree } n \text{ or less with coefficients in } \mathbb{F}\}$
$F(D) = \{f : f \text{ is a function from } D \text{ to } \mathbb{R}\}$
$C(D) = \{f : f \text{ is a continuous function from } D \text{ to } \mathbb{R}\}$

Exercises

1. Verify that the set, with its well known operations of addition and scalar multiplication, forms a vector space.
   (a) $\mathcal{M}_{m\times n}(\mathbb{R})$
   (b) $F([0, 1])$ [$\mathbb{S}$]-303
   (c) $C^3(\mathbb{R})$
   (d) $\mathbb{R}^\mathbb{N}$ [The "well known" operations are done component-wise.]
2. Verify that S is a subspace of the vector space V. You do not have to verify that V is a vector space.
   (a) S = the x-axis $= \{(x, y) : y = 0\}$; $V = \mathbb{R}^2$ [$\mathbb{S}$]-303
   (b) S = the line with slope m passing through the origin $= \{(x, y) : y = mx\}$; $V = \mathbb{R}^2$
   (c) S = the set of polynomials of degree two or less having 12 as a root $= \{p(x) = ax^2+bx+c : p(12) = 0\}$; $V = \mathbb{P}_2(\mathbb{R})$
   (d) S = the set of cubic polynomials with degree at most three and 3 and 18 as roots $= \{q(x) = a_0 + a_1x + a_2x^2 + a_3x^3 : q(3) = q(18) = 0\}$; $V = \mathbb{P}_3(\mathbb{R})$ [$\mathbb{S}$]-304
   (e) S = the set of functions on $[0, 1]$ whose graphs pass through $(0, 0)$ and $(1, 0)$; $V = F([0, 1])$
   (f) S = the set of real number sequences that converge

[2] $\mathbb{F}$ could actually be any field.

to zero $= \{r_1, r_2, \ldots : \lim_{n\to\infty} r_n = 0\}$; $V = \mathbb{R}^{\mathbb{N}}$

(g) S = the set of real 3×3 matrices with zero determinant; $\mathcal{M}_{3\times 3}(\mathbb{R})$

3. Show that S is not a subspace of the vector space V. You do not have to verify that V is a vector space.

(a) S = the first quadrant of the plane, including the axes $= \{(x,y) : x \geq 0 \text{ and } y \geq 0\}$; $V = \mathbb{R}^2$ [𝕊]-304

(b) S = the first and third quadrants of the plane, including the axes $= \{(x,y) : x \geq 0 \text{ and } y \geq 0\} \cup \{(x,y) : x \leq 0 \text{ and } y \leq 0\}$; $V = \mathbb{R}^2$

(c) S = the closed disk of radius 5 with center at the origin $= \{(x,y) : x^2 + y^2 \leq 25\}$; $V = \mathbb{R}^2$

(d) S = the set of polynomials of degree two or less with y-intercept $(0,-3) = \{p(x) = ax^2 + bx + c : p(0) = -3\}$; $V = \mathbb{P}_2(\mathbb{R})$ [𝕊]-304

(e) S = the set of polynomials of degree three or less such that 7 is not a root $= \{q(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 : q(7) \neq 0\}$; $V = \mathbb{P}_3(\mathbb{R})$

(f) S = the set of polynomials of degree four or less with its real roots between 0 and 10 union the set containing the zero polynomial $= \{q(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a^4 x^4 : q(x_0) = 0 \Rightarrow 0 \leq x_0 \leq 10 \text{ or } q(x) = 0 \text{ for all } x\}$; $V = \mathbb{P}_4(\mathbb{R})$

(g) S = the set of all real number sequences that converge to three $= \{r_1, r_2, \ldots : \lim_{n\to\infty} r_n = 3\}$; $V = \mathbb{R}^{\mathbb{N}}$

(h) S = the set of all real 3×3 matrices with determinant one; $\mathcal{M}_{3\times 3}(\mathbb{R})$

4. Describe span$S$ geometrically.

(a) $S = \left\{ \begin{bmatrix} 5 \\ 1 \end{bmatrix} \right\}$ [𝕊]-304

(b) $S = \left\{ \begin{bmatrix} -1 \\ 3 \end{bmatrix} \right\}$

(c) $S = \left\{ \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix} \right\}$ [𝕊]-304

(d) $S = \left\{ \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ -6 \end{bmatrix} \right\}$ [𝔸]-352

(e) $S = \left\{ \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \begin{bmatrix} -10 \\ 2 \end{bmatrix} \right\}$

(f) $S = \left\{ \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 7 \end{bmatrix} \right\}$

(g) $S = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}$

(h) $S = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$

(i) $S = \left\{ \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ [𝔸]-352

(j) $S = \left\{ \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \right\}$

5. Describe span$S$ in words.

(a) $S = \{\langle 1,1,2,1,1,1,1,1,1,1,\ldots\rangle\}$ [𝔸]-352

(b) $S = \{\langle 1,2,3,4,5,6,7,8,9,10,\ldots\rangle\}$

(c) $S = \{p(x) = x(1-x)\}$

(d) $S = \{p(x) = 3\}$ [𝔸]-352

(e) $S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$

(f) $S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} \right\}$ [𝕊]-305

6. Let S be the set of all vectors of the form $\begin{bmatrix} s \\ 2s \end{bmatrix}$. Find a set of vectors in $\mathbb{R}^2$ whose span is S.

7. Let S be the set of all vectors of the form $\begin{bmatrix} 0 \\ -3t \\ 5t \end{bmatrix}$. Find a set of vectors in $\mathbb{R}^3$ whose span is S.

8. Let S be the set of all vectors of the form $\begin{bmatrix} 3t \\ -t \\ 5t \end{bmatrix}$. Find a set of vectors in $\mathbb{R}^3$ whose span is S. [𝕊]-305

9. Let S be the set of all vectors of the form $\begin{bmatrix} 3t-2s \\ -t+s \\ 5t+s \end{bmatrix}$. Find a set of vectors in $\mathbb{R}^3$ whose span is S. [𝕊]-305

10. Let S be the set of all vectors of the form $\begin{bmatrix} 3t-4r \\ -t+4r \\ 5t \end{bmatrix}$. Find a set of vectors in $\mathbb{R}^3$ whose span is S.

11. Let S be the set of all vectors of the form $\begin{bmatrix} s+t \\ 2s-t \end{bmatrix}$. Find a set of two vectors in $\mathbb{R}^2$ whose span is S. Is $\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} = S$?

12. Is $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ in span$S$?

(a) $S = \left\{ \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \right\}$

(b) $S = \left\{ \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}, \begin{bmatrix} 10 \\ 2 \\ -6 \end{bmatrix} \right\}$ [𝕊]-305

(c) $S = \left\{ \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix}, \begin{bmatrix} -10 \\ -24 \\ 41 \end{bmatrix} \right\}$ [𝔸]-352

(d) $S = \left\{ \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix}, \begin{bmatrix} -10 \\ -24 \\ 41 \end{bmatrix}, \begin{bmatrix} 8 \\ 1 \\ -2 \end{bmatrix} \right\}$

(e) $S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$

(f) $S = \left\{ \begin{bmatrix} -\frac{2}{7} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{129}{1111} \\ -3 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{122}{37} \\ -\frac{37}{21} \\ \frac{1}{8} \end{bmatrix} \right\}$ [$\mathbb{A}$]-352

13. Is $\mathbf{v}$ in span $\left\{\sin^2\theta, \cos^2\theta\right\}$?

(a) $\mathbf{v} = (f(\theta) = 3)$

(b) $\mathbf{v} = (f(\theta) = \sin(2\theta))$

(c) $\mathbf{v} = (f(\theta) = \cos(2\theta))$ [$\mathbb{S}$]-305

(d) $\mathbf{v} = \left(f(\theta) = \sin^2(\theta/2)\right)$

14. Is $\mathbf{v}$ in span $\{\langle 1,2,3,4,5,\ldots\rangle, \langle 0,4,0,0,0,\ldots\rangle, \langle 1,1,1,1,1,\ldots\rangle\}$?

(a) $\mathbf{v} = \langle 5, 10, 15, 20, 25, \ldots\rangle$ [$\mathbb{A}$]-352

(b) $\mathbf{v} = \langle 3, 87, 9, 12, 15, \ldots\rangle$

(c) $\mathbf{v} = \langle 78, 81, 84, 87, 90, \ldots\rangle$ [$\mathbb{S}$]-305

(d) $\mathbf{v} = \langle 2, 4, 8, 16, 32, \ldots\rangle$

(e) $\mathbf{v} = \left\langle -\frac{1}{2}, -1, -\frac{3}{2}, -2, -\frac{5}{2}, \ldots\right\rangle$

(f) $\mathbf{v} = \langle 0, 32, 4, 6, 8, \ldots\rangle$ [$\mathbb{A}$]-352

15. Is $\mathbf{b}$ in the span of the columns of M?

(a) SageMathCell 44 $\mathbf{b} = \begin{bmatrix} -240 \\ -406 \\ -416 \end{bmatrix}$;

$M = \begin{bmatrix} -499 & -288 & 232 \\ -425 & -161 & 125 \\ 306 & -348 & 141 \end{bmatrix}$ [$\mathbb{S}$]-306

(b) SageMathCell 45 $\mathbf{b} = \begin{bmatrix} -284 \\ 146 \\ -187 \end{bmatrix}$;

$M = \begin{bmatrix} 237 & -228 & -234 \\ -440 & -772 & 36 \\ 243 & -612 & -366 \end{bmatrix}$

(c) SageMathCell 46 $\mathbf{b} = \begin{bmatrix} 486 \\ -352 \\ 418 \\ 273 \end{bmatrix}$;

$M = \begin{bmatrix} -118 & 35 & -94 & -170 \\ 277 & -101 & 496 & -336 \\ 332 & 25 & 393 & 164 \\ 182 & -233 & 326 & -252 \end{bmatrix}$ [$\mathbb{A}$]-352

(d) SageMathCell 47 $\mathbf{b} = \begin{bmatrix} -93 \\ 130 \\ -120 \\ -437 \end{bmatrix}$;

$M = \begin{bmatrix} -133 & -77 & -153 & 209 \\ -69 & 323 & 490 & -98 \\ -419 & 15 & 305 & 129 \\ -157 & -10 & -230 & 377 \end{bmatrix}$

16. Is $\mathbf{v}$ in span $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$?

(a) SageMathCell 48 $\mathbf{v} = \begin{bmatrix} -7424 \\ 3075 \\ -1060 \end{bmatrix}$;

$\mathbf{u}_1 = \begin{bmatrix} -12 \\ -20 \\ 20 \end{bmatrix}$; $\mathbf{u}_2 = \begin{bmatrix} -560 \\ 255 \\ -100 \end{bmatrix}$;

$\mathbf{u}_3 = \begin{bmatrix} -8 \\ 25 \\ -20 \end{bmatrix}$; [$\mathbb{A}$]-352

(b) SageMathCell 49 $\mathbf{v} = \begin{bmatrix} 411 \\ -166 \\ -258 \end{bmatrix}$;

$\mathbf{u}_1 = \begin{bmatrix} 20 \\ 11 \\ -5 \end{bmatrix}$; $\mathbf{u}_2 = \begin{bmatrix} -23 \\ -6 \\ 17 \end{bmatrix}$; $\mathbf{u}_3 = \begin{bmatrix} -193 \\ 100 \\ 397 \end{bmatrix}$;

(c) SageMathCell 50 $\mathbf{v} = \begin{bmatrix} 2 \\ -3 \\ 18 \\ 44 \end{bmatrix}$;

$\mathbf{u}_1 = \begin{bmatrix} 10 \\ 11 \\ 1 \\ -7 \end{bmatrix}$; $\mathbf{u}_2 = \begin{bmatrix} -4 \\ 4 \\ -9 \\ -7 \end{bmatrix}$; $\mathbf{u}_3 = \begin{bmatrix} 8 \\ -12 \\ -5 \\ -6 \end{bmatrix}$; [$\mathbb{A}$]-352

(d) SageMathCell 51 $\mathbf{v} = \begin{bmatrix} 79 \\ 33 \\ -312 \\ 366 \end{bmatrix}$;

$\mathbf{u}_1 = \begin{bmatrix} 4 \\ -2 \\ -12 \\ 11 \end{bmatrix}$; $\mathbf{u}_2 = \begin{bmatrix} -10 \\ -9 \\ 9 \\ -6 \end{bmatrix}$; $\mathbf{u}_3 = \begin{bmatrix} -3 \\ -8 \\ 7 \\ -11 \end{bmatrix}$;

17. For each pair $\mathbf{v}$ and S, $\mathbf{v}$ is in span$S$. Show that $S \cup \{\mathbf{v}\}$ is linearly dependent.

(a) $\mathbf{v} = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$; $S = \left\{ \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \end{bmatrix} \right\}$ [$\mathbb{A}$]-352

(b) $\mathbf{v} = \begin{bmatrix} 18 \\ 21 \end{bmatrix}$; $S = \left\{ \begin{bmatrix} 12 \\ -3 \end{bmatrix}, \begin{bmatrix} 6 \\ 7 \end{bmatrix}, \begin{bmatrix} 8 \\ -1 \end{bmatrix} \right\}$

(c) $\mathbf{v} = 5t^2 - 9t + 5$; $S = \left\{1 + t^2, 3t\right\}$

(d) $\mathbf{v} = 5t^2 - 9t + 5$; $S = \left\{1, t, t^2\right\}$ [$\mathbb{A}$]-352

(e) $\mathbf{v} = \langle 0, 0, 0, 0, 0, \ldots\rangle$; $S = \{\langle 2, -3, 4, 5, 1, \ldots\rangle\}$

(f) $\mathbf{v} = \langle 2, 3, 4, 6, 6, 9, 8, 12, \ldots\rangle$; $S = \{\langle 1, 0, 2, 0, 3, 0, 4, 0, \ldots\rangle, \langle 0, 1, 0, 2, 0, 3, 0, 4, \ldots\rangle\}$

18. Argue that (in general) if $\mathbf{v}$ is in span$S$, then $S \cup \{\mathbf{v}\}$ is linearly dependent.

19. Find k so that $kt^2 + 9t - 8$ is in span $\left\{3t^2 - 4, 4t^2 - 3t\right\}$.

20. Let $M = \begin{bmatrix} 2 & -3 & 5 \\ 1 & 8 & 2 \end{bmatrix}$. Is the set of all $\mathbf{v}$ such that $M\mathbf{v} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ a subspace of $\mathbb{R}^2$? of $\mathbb{R}^3$?

21. Suppose $\mathbf{u}_1$ and $\mathbf{u}_2$ both have the property of being a zero vector. That is, for any vector $\mathbf{v}$, $\mathbf{u}_1 + \mathbf{v} = \mathbf{v}$ and $\mathbf{u}_2 + \mathbf{v} = \mathbf{v}$.

Supply a reason for each line of the string of equalities proving that $\mathbf{u}_1 = \mathbf{u}_2$, thus proving that the zero vector is unique [and therefore $\mathbf{0}$ is ***the*** additive identity]. [𝔸]-352

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{u}_2 + \mathbf{u}_1 \\ &= \mathbf{u}_1 + \mathbf{u}_2 \\ &= \mathbf{u}_2 \end{aligned}$$

22. Prove that for any vector $\mathbf{v}$ in a vector space, $-1 \cdot \mathbf{v}$ is an additive inverse of $\mathbf{v}$. Use the fact that $0\mathbf{v} = \mathbf{0}$.
23. Suppose $\mathbf{u}_1$ and $\mathbf{u}_2$ both have the property of being an additive inverse of $\mathbf{v}$. That is, for this arbitrary but particular vector $\mathbf{v}$, $\mathbf{u}_1 + \mathbf{v} = \mathbf{0}$ and $\mathbf{u}_2 + \mathbf{v} = \mathbf{0}$. Supply a reason for each line of the string of equalities proving that $\mathbf{u}_1 = \mathbf{u}_2$, thus proving that additive inverses are unique [and therefore $-1 \cdot \mathbf{v}$ is ***the*** additive inverse of $\mathbf{v}$].

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{u}_1 + \mathbf{0} \\ &= \mathbf{u}_1 + (\mathbf{u}_2 + \mathbf{v}) \\ &= \mathbf{u}_1 + (\mathbf{v} + \mathbf{u}_2) \\ &= (\mathbf{u}_1 + \mathbf{v}) + \mathbf{u}_2 \\ &= \mathbf{0} + \mathbf{u}_2 \\ &= \mathbf{u}_2 \end{aligned}$$

24. Let c be any scalar. Supply a reason for each line of the string of equalities proving that $c\mathbf{0} = \mathbf{0}$.

$$\begin{aligned} c\mathbf{0} &= c\mathbf{0} + \mathbf{0} \\ &= c\mathbf{0} + (c\mathbf{0} + (-(c\mathbf{0}))) \\ &= (c\mathbf{0} + c\mathbf{0}) + (-(c\mathbf{0})) \\ &= c(\mathbf{0} + \mathbf{0}) + (-(c\mathbf{0})) \\ &= c\mathbf{0} + (-(c\mathbf{0})) \\ &= \mathbf{0} \end{aligned}$$

25. Suppose $c\mathbf{u} = \mathbf{0}$ for some nonzero scalar c. Create a string of equalities showing the $\mathbf{u} = \mathbf{0}$ and justify each equality in the string.
26. Suppose V is a vector space, H is a subset of V, and S is any subspace of V containing H. Explain why span$H$ must be a subset of S. This shows that span$H$ is the smallest subspace of V containing H.
27. Let S and T be subspaces of a vector space V. Show that $S \cap T$ is a subspace of V.
28. Find subspaces S and T of a vector space V such that $S \cup T$ is not a subspace of V.

Answers

not closed With $T = \{t, t^2, 1 + t^2\}$, $t \in T$ and $t^2 \in T$, but $t + t^2$, the sum of two elements in T, is not itself in T. $17t^2$, a scalar multiple of an element of T, is not itself in T.

closure of linear combinations With $S = \{\alpha t + \beta t^2 + \gamma(1 + t^2) : \alpha, \beta, \gamma \in \mathbb{R}\}$, an arbitrary element of S has the form $\alpha t + \beta t^2 + \gamma(1 + t^2)$ for some scalars α, β, γ. The scalar multiple of an arbitrary element of S, $r(\alpha t + \beta t^2 + \gamma(1 + t^2)) = (r\alpha)t + (r\beta)t^2 + (r\gamma)(1 + t^2)$ has the form of an element of S and is therefore in S. The sum of two elements of S, $\left[\alpha_1 t + \beta_1 t^2 + \gamma_1(1 + t^2)\right] + \left[\alpha_2 t + \beta_2 t^2 + \gamma_2(1 + t^2)\right] = (\alpha_1 + \alpha_2)t + (\beta_1 + \beta_2)t^2 + (\gamma_1 + \gamma_2)(1 + t^2)$ has the form of an element of S and is therefore in S too.

the zero vector Proving that $0\mathbf{v} = \mathbf{0}$ for any vector $\mathbf{v}$ of any vector space V is a very abstract and subtle chore. Where to start is a particularly befuddling question. Proofs of this nature are very difficult when seen for the first time. You are in good company if you did not come up with a proof of your own. One way to proceed is to start with one side and produce a list of equalities that flow logically from the properties and lead to the other side. Let $\mathbf{v}$ be a particular but arbitrary element of a vector space V.

$$\begin{aligned} 0\mathbf{v} &= 0\mathbf{v} + \mathbf{0} && [\mathbf{0} \text{ is the additive identity}] \\ &= 0\mathbf{v} + (0\mathbf{v} + (-0\mathbf{v})) && [\text{property of additive inverses}] \\ &= (0\mathbf{v} + 0\mathbf{v}) + (-0\mathbf{v}) && [\text{addition is associative}] \\ &= (0 + 0)\mathbf{v} + (-0\mathbf{v}) && [\text{scalars distribute over vectors}] \\ &= 0\mathbf{v} + (-0\mathbf{v}) && [\text{substitution of } 0 \text{ for } 0 + 0] \\ &= \mathbf{0} && [\text{property of additive inverses}] \end{aligned}$$

span is a subspace Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be elements of a vector space and set $H = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$. Then

1. [property 4] $0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_k = \mathbf{0} + \mathbf{0} + \cdots + \mathbf{0} = \mathbf{0}$ is in H.
2. [property 1] for any elements $\mathbf{u}$ and $\mathbf{v}$ of H, there are scalars $b_1, b_2, \ldots, b_k$ and $c_1, c_2, \ldots, c_k$ such that

$$\mathbf{u} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k \quad \text{and} \quad \mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

so

$$\begin{aligned}\mathbf{u}+\mathbf{v} &= (b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k) + (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k) \\ &= b_1\mathbf{v}_1 + c_1\mathbf{v}_1 + b_2\mathbf{v}_2 + +c_2\mathbf{v}_2 \cdots + b_k\mathbf{v}_k + c_k\mathbf{v}_k \\ &= (b_1+c_1)\mathbf{v}_1 + (b_2+c_2)\mathbf{v}_2 + \cdots + (b_k+c_k)\mathbf{v}_k\end{aligned}$$

is in H.

3. [property 6] for any element $\mathbf{u}$ of H, there are scalars $b_1, b_2, \ldots, b_k$ such that

$$\mathbf{u} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k$$

so given any scalar s,

$$\begin{aligned}s\mathbf{u} &= s\,(b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k) \\ &= s(b_1\mathbf{v}_1) + s(b_2\mathbf{v}_2) + \cdots + s(b_k\mathbf{v}_k) \\ &= (sb_1)\mathbf{v}_1 + (sb_2)\mathbf{v}_2 + \cdots + (sb_k)\mathbf{v}_k\end{aligned}$$

is in H.

4.2 Basis and Dimension

Every vector in $\mathbb{R}^n$ can be written as a unique linear combination of the columns of the $n \times n$ identity matrix. Think about it for a moment. Can you justify this claim? Answer on page 129.

Taking the perspective that the matrix-column product $I\mathbf{v}$ is a linear combination of the columns of I with coefficients from $\mathbf{v}$, what we are claiming is that for any vector $\mathbf{b}$, the equation

$$I\mathbf{v} = \mathbf{b} \tag{4.2.1}$$

has exactly one solution $\mathbf{v}$. By the invertible matrix theorem, this is equivalent to claiming that I is invertible, which of course is true!

Noting that the solution of (4.2.1) is $\mathbf{v} = \mathbf{b}$ we see the vector $\mathbf{b}$ itself holds the coefficients of the proclaimed linear combination. There is nothing ground-breaking about the calculation itself. However, it strikes at the essence of elements of $\mathbb{R}^n$, paving the way for abstraction to arbitrary vector spaces.

Given any subset of a vector space, we can represent linear combinations of these vectors by the element of $\mathbb{R}^n$ holding the coefficients of the linear combination. If there were special subsets where each element of the vector space could be represented by a unique linear combination (as in the example of the columns of I), this perspective would hold promise. As it turns out, there are many such sets. For example, $T = \{t, t^2, 1 + t^2\}$ forms such a set in $\mathbb{P}_2(\mathbb{R})$. Can you verify this? Answer on page 129.

In $\mathbb{R}^n$ the columns of any invertible matrix M form such a set. If M is invertible, the invertible matrix theorem tells us that for any vector $\mathbf{b}$ there is exactly one solution $\mathbf{v}$ of the equation

$$M\mathbf{v} = \mathbf{b}.$$

In terms of linear combinations, every vector $\mathbf{b}$ can be written as a unique linear combination of the columns of M. In the terminology of the previous section, the fact that $M\mathbf{v} = \mathbf{b}$ always has a solution is to say that the columns of M span $\mathbb{R}^n$. The linear independence of the columns of M ensures uniqueness of solution (by theorem 5). These are the characteristics of the columns of an invertible matrix that makes it a special set—linear independence and span.

The terms *linear independence* and *span* have meaning in *any* vector space, not just $\mathbb{R}^n$, motivating the following definitions. A subset S of a vector space V is called a **spanning set** if span$S = V$. A subset of a vector space V is called a **basis** (of V) if it is a linearly independent spanning set.

From the discussion that led to the definition of a basis, we can be sure that a basis of $\mathbb{R}^n$ has the special property that every vector in $\mathbb{R}^n$ can be written as a unique linear combination of the vectors in the basis. However, theorem 5, which provided uniqueness does not apply to arbitrary vector spaces. We will need to argue that if $\mathcal{B}$ is a basis of an arbitrary vector space V and $\mathbf{v}$ is in V, then $\mathbf{v}$ is expressible as a unique linear combination of the vectors in $\mathcal{B}$.

Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis (linearly independent spanning set) for a vector space V and let $\mathbf{v}$ be a vector in V. Because $\mathcal{B}$ is a spanning set, every vector in V can be written as a linear combination of the elements of $\mathcal{B}$, including $\mathbf{v}$. Thus $\mathbf{v}$ is expressible as at least one linear combination of the vectors in $\mathcal{B}$. It remains to show $\mathbf{v}$ is expressible as at most one linear combination of the vectors in $\mathcal{B}$. To that end, suppose $\mathbf{v}$ has two representations as linear combinations of the elements of $\mathcal{B}$. That is,

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n \quad \text{and} \quad \mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_n\mathbf{v}_n.$$

Then $\mathbf{0} = \mathbf{v} - \mathbf{v} = (a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n) - (b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_n\mathbf{v}_n) = (a_1 - b_1)\mathbf{v}_1 + (a_2 - b_2)\mathbf{v}_2 + \cdots + (a_n - b_n)\mathbf{v}_n$. Since $\mathcal{B}$ is a linearly independent set, the only solution of this equation is

$$a_1 - b_1 = a_2 - b_2 = \cdots = a_n - b_n = 0$$

and therefore the two linear combinations are equal. So, given a basis of a vector space, every element of the vector space can be written as a unique linear combination of the elements of the basis.

Bases (the plural of basis) have another important property. If there is a basis of a vector space with n elements, then any subset of the vector space with more than n elements is linearly dependent.

Suppose $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of a vector space V and $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ where $p > n$. Since $\mathcal{B}$ is a spanning set, each element of S can be written as a linear combination of the elements of $\mathcal{B}$. Let

$$\begin{aligned}
\mathbf{u}_1 &= M_{1,1}\mathbf{v}_1 + M_{2,1}\mathbf{v}_2 + \cdots + M_{n,1}\mathbf{v}_n \\
\mathbf{u}_2 &= M_{1,2}\mathbf{v}_1 + M_{2,2}\mathbf{v}_2 + \cdots + M_{n,2}\mathbf{v}_n \\
&\vdots \\
\mathbf{u}_p &= M_{1,p}\mathbf{v}_1 + M_{2,p}\mathbf{v}_2 + \cdots + M_{n,p}\mathbf{v}_n
\end{aligned}$$

Then

$$M = \begin{bmatrix} M_{1,1} & M_{1,2} & \cdots & M_{1,p} \\ M_{2,1} & M_{2,2} & \cdots & M_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ M_{n,1} & M_{n,2} & \cdots & M_{n,p} \end{bmatrix}$$

has more columns than rows, so M cannot have a pivot in each column. By theorem 5, there is a nonzero vector $\mathbf{w}$ such that $M\mathbf{w} = \mathbf{0}$. Let $\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_p \end{bmatrix}^T$ be such a vector. Then

$$\begin{aligned}
w_1\mathbf{u}_1 + w_2\mathbf{u}_2 + \cdots + w_p\mathbf{u}_p = &\quad w_1 M_{1,1}\mathbf{v}_1 + w_1 M_{2,1}\mathbf{v}_2 + \cdots + w_1 M_{n,1}\mathbf{v}_n \\
&+ w_2 M_{1,2}\mathbf{v}_1 + w_2 M_{2,2}\mathbf{v}_2 + \cdots + w_2 M_{n,2}\mathbf{v}_n \\
&\vdots \\
&+ w_p M_{1,p}\mathbf{v}_1 + w_p M_{2,p}\mathbf{v}_2 + \cdots + w_p M_{n,p}\mathbf{v}_n \\
= &\quad (M_{1,1}w_1 + M_{1,2}w_2 + \cdots + M_{1,p}w_p)\mathbf{v}_1 \\
&+ (M_{2,1}w_1 + M_{2,2}w_2 + \cdots + M_{2,p}w_p)\mathbf{v}_2 \\
&\vdots \\
&+ (M_{n,1}w_1 + M_{n,2}w_2 + \cdots + M_{n,p}w_p)\mathbf{v}_n \\
= &\ 0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_n \\
= &\ \mathbf{0}.
\end{aligned}$$

Since $\mathbf{w} \neq \mathbf{0}$, we have demonstrated a nontrivial linear combination of the vectors in S that sum to $\mathbf{0}$, showing that S is linearly dependent. This fact is important enough to repeat as a theorem.

Theorem 9. *If a vector space V has a basis with n elements, then any subset of V containing more than n elements is linearly dependent.*

To understand the implications of this theorem, let V be a vector space with basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and consider two subsets of V: F containing fewer than n vectors and G containing greater than n vectors. G cannot be a basis because it has more than n elements (which makes it a linearly dependent set). F cannot be a basis for V because if it were, that would make $\mathcal{B}$ linearly dependent thereby contradicting the fact that $\mathcal{B}$ is a basis. In other words, if a vector space V admits a basis with n elements, all bases of V have n elements. This number n is thus a characteristic of V and deserves a name. We call the number of elements in a basis the **dimension** of a vector space. The **trivial vector space**, $\{\mathbf{0}\}$ contains only the one vector $\mathbf{0}$ and, by definition, has dimension 0. Observe that

1. the dimension of $\mathbb{R}^n$ is n, and

2. the dimension of $\mathbb{P}_2(\mathbb{R})$ is 3

because

1. the columns of $I_{n\times n}$ form a basis of $\mathbb{R}^n$, and

2. $T = \{t, t^2, 1 + t^2\}$ forms a basis of $\mathbb{P}_2(\mathbb{R})$.

$\{I_{:,1}, I_{:,2}, \ldots, I_{:,n}\}$ is called the **standard basis** of $\mathbb{R}^n$ and $\{1, t, \ldots, t^n\}$ is called the **standard basis** of $\mathbb{P}_n(\mathbb{R})$.

We already know that if a vector space V has dimension n, then any subset of V containing fewer than n elements is not a basis. It therefore must either be linearly dependent or not span V. We will argue that such a set does not span V. The statement is clearly true when the dimension of V is 0 (there are no sets with less than 0 elements) and $n = 1$ (only the empty set has less than one element, but the empty set does not span a nonempty vector space). For $n \geq 2$, let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ be a subset of V with $p < n$ and suppose S is a spanning set. Because S spans V but is not a basis, S must be linearly dependent. This is a contradiction in the case $n = 2$ since any set with one element is linearly independent. In the case $n > 2$, S contains some element that can be written as a linear combination of the others. By rearranging the order of the elements in S if necessary, let $\mathbf{u}_1$ be the element that can be written as a linear combination of the others and write

$$\mathbf{u}_1 = c_2\mathbf{u}_2 + c_3\mathbf{u}_3 + \cdots c_p\mathbf{u}_p. \tag{4.2.2}$$

Note that $\hat{S} = S \setminus \{\mathbf{u}_1\} = \{\mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_p\}$ still spans V: for arbitrary $\mathbf{v}$ in V, write

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots c_p\mathbf{u}_p \tag{4.2.3}$$

and substitute (4.2.2) into (4.2.3), thereby writing $\mathbf{v}$ as a linear combination of the elements of $\hat{S}$. If $\hat{S}$ is linearly dependent, repeat the process of removing an element that can be written as a linear combination of the others, and continue until the remaining set is linearly independent (and still spanning). This will happen when there is one element left, if not sooner, so the process is guaranteed to end. At this point what is left is a basis with less than n elements, an impossibility. It must be that S is not spanning.

Key Concepts

spanning set S is a spanning set of a vector space V if span$S = V$.

basis a linearly independent spanning set of a vector space.

dimension the number of vectors in a basis of a vector space.

standard basis of $\mathbb{R}^n$ the columns of $I_{n\times n}$.

standard basis of $\mathbb{P}_n(\mathbb{R})$ $\{1, t, \ldots, t^n\}$.

linear dependence if a vector space V has dimension n, any subset of V with more than n elements is linearly dependent.

spanning if a vector space V has dimension n, any subset of V with less than n elements does not span V.

bases all bases of a given vector space have the same number of elements.

trivial vector space $\{\mathbf{0}\}$ contains only the one vector $\mathbf{0}$ and, by definition, has dimension 0.

Exercises

1. State the dimension of the vector space.

   (a) $\mathbb{R}^2$

   (b) $\mathbb{R}^5$ [A]-352

   (c) $\mathbb{R}^{12}$

   (d) $\mathbb{R}^n$

   (e) $\mathbb{P}_2(\mathbb{R})$

   (f) $\mathbb{P}_5(\mathbb{R})$ [A]-352

   (g) $\mathbb{P}_{12}(\mathbb{R})$

   (h) $\mathbb{P}_n(\mathbb{R})$

2. Explain why the set is not a basis of $\mathbb{R}^3$.

   (a) $\left\{\begin{bmatrix} -321 \\ 238 \\ -250 \end{bmatrix}, \begin{bmatrix} -15 \\ -298 \\ -64 \end{bmatrix}\right\}$

   (b) $\left\{\begin{bmatrix} 384 \\ -391 \\ -339 \end{bmatrix}, \begin{bmatrix} -78 \\ -262 \\ 349 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right\}$

   (c) $\left\{\begin{bmatrix} -91 \\ 412 \\ 0 \end{bmatrix}, \begin{bmatrix} 147 \\ 198 \\ 0 \end{bmatrix}, \begin{bmatrix} 423 \\ -218 \\ 0 \end{bmatrix}\right\}$ [A]-352

   (d) $\left\{\begin{bmatrix} -51 \\ 165 \\ -136 \\ -34 \end{bmatrix}, \begin{bmatrix} 0 \\ 414 \\ 308 \\ -264 \end{bmatrix}, \begin{bmatrix} 491 \\ -37 \\ 198 \\ 36 \end{bmatrix}\right\}$

(e) $\left\{\begin{bmatrix}115\\136\\111\end{bmatrix},\begin{bmatrix}46\\90\\-131\end{bmatrix},\begin{bmatrix}-101\\-122\\105\end{bmatrix},\begin{bmatrix}-94\\-91\\-148\end{bmatrix}\right\}$

3. Explain why the set is not a basis of $\mathbb{P}_3(\mathbb{R})$.

   (a) $\{168+217t-115t^2+383t^3, 60+349t+498t^2+243t^3\}$

   (b) $\{-407+342t-94t^2-354t^3, 188-272t^2-18t^3, 117+114t-437t^2+493t^3\}$

   (c) $\{-9+t, -4+9t-t^2, 7-2t^2, -2+7t-6t^2\}$

   (d) $\{0, t, t^2, t^3\}$

   (e) $\{t, t^2, t^3, t^4\}$ [A]-352

   (f) $\{27-70t, -89t^2+52t^3, 1+58t, -27t^2-74t^3, 28-100t\}$

4. Each set is a subset of a vector space. In each set, one vector is the sum of the other two. (i) Name a vector space containing the set, and (ii) find a proper subset with the same span.

   (a) $\left\{\begin{bmatrix}9\\12\\-7\\18\end{bmatrix},\begin{bmatrix}18\\8\\0\\7\end{bmatrix},\begin{bmatrix}9\\-4\\7\\-11\end{bmatrix}\right\}$

   (b) $\{9-4t+7t^2, -20+8t+17t^2, -11+4t+24t^2\}$ [A]-352

   (c) $\{\langle -4,-7,-10,-13,\ldots\rangle, \langle 1,2,3,4,\ldots\rangle, \langle -5,-9,-13,-17,\ldots\rangle\}$

   (d) $\left\{\begin{bmatrix}-4&-19\\14&-13\end{bmatrix},\begin{bmatrix}-19&-2\\14&-1\end{bmatrix},\begin{bmatrix}-23&-21\\28&-14\end{bmatrix}\right\}$ [S]-306

   (e) $\{3\sin\theta-2\cos\theta, 8\sin\theta+2\cos\theta, 5\sin\theta+4\cos\theta\}$

5. Redo part (ii) of exercise 4, this time stating a different subset with the same span. Is this subset a basis for the span of the set? [S]-306 [A]-352

6. Find the dimension of the span of the set.

   (a) $\left\{\begin{bmatrix}1\\1\\2\\2\end{bmatrix},\begin{bmatrix}2\\2\\4\\4\end{bmatrix},\begin{bmatrix}5\\5\\10\\10\end{bmatrix}\right\}$

   (b) $\left\{3+2t^2, 1, 1-t^2\right\}$

   (c) $\left\{1+2t, 3+t-2t^2, -5+4t^2, -5t-2t^2\right\}$ [S]-306

   (d) $\left\{\begin{bmatrix}1&0\\0&0\end{bmatrix},\begin{bmatrix}0&1\\0&0\end{bmatrix},\begin{bmatrix}0&0\\1&0\end{bmatrix},\begin{bmatrix}0&0\\0&1\end{bmatrix}\right\}$

   (e) $\left\{\cos^2\theta, \sin^2\theta, 1\right\}$ [A]-352

   (f) $\left\{\begin{bmatrix}1\\2\\3\end{bmatrix},\begin{bmatrix}3\\6\\9\end{bmatrix},\begin{bmatrix}-2\\2\\-1\end{bmatrix},\begin{bmatrix}4\\-4\\2\end{bmatrix}\right\}$ [A]-352

7. Justify the claim that $\{1, t, t^2, \ldots, t^n\}$ is a basis of $\mathbb{P}_n(\mathbb{R})$.

8. Suppose $\operatorname{span}\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\} = \mathbb{P}_3(\mathbb{R})$. Explain why $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$ is a basis of $\mathbb{P}_3(\mathbb{R})$. [A]-352

9. Suppose $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$ is a linearly independent set in $\mathbb{R}^4$. Explain why $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$ is a basis of $\mathbb{R}^4$. [A]-353

10. Let V be a vector space with dimension n. Explain why a spanning set with n elements must be a basis.

11. Let V be a vector space with dimension n. Explain why a linearly independent set with n elements must be a basis.

12. Do the columns of the matrix (considered as a set of vectors) form a basis for $\mathbb{R}^4$?

    (a) $\begin{bmatrix}-11&8&18&15\\19&2&5&-3\\17&11&20&-2\end{bmatrix}$

    (b) $\begin{bmatrix}-13&19&19&16\\-17&-16&-16&14\\-19&5&5&-10\\-14&15&15&4\end{bmatrix}$ [S]-307

    (c) $\begin{bmatrix}-14&14&8\\-2&18&-17\\-16&7&-6\\12&-18&-4\end{bmatrix}$

    (d) $\begin{bmatrix}-17&18&-12&9\\0&-8&4&17\\0&0&12&-9\\0&0&0&-10\end{bmatrix}$ [A]-353

    (e) $\begin{bmatrix}-7&18&19&-12\\0&0&8&2\\0&0&0&-1\\0&0&0&0\end{bmatrix}$

    (f) $\begin{bmatrix}9&3&13&11&-5&-6\\-17&12&-14&10&-2&-7\\-13&8&0&16&-15&18\\-8&19&-12&15&2&1\end{bmatrix}$ [A]-353

13. Do the columns of the matrix (considered as a set of vectors) form a basis for $\mathbb{R}^6$?

    (a) SageMathCell 52

    $\begin{bmatrix}-89&86&-47&69&-88&-61\\-27&27&95&-16&93&133\\-50&150&-73&52&-17&24\\-6&78&60&84&41&91\\132&-110&-126&79&-90&-137\\96&-124&-41&147&75&-117\end{bmatrix}$

    (b) SageMathCell 53

    $\begin{bmatrix}7&-4&59&-15&49&56\\21&-17&232&-55&197&242\\-21&9&-145&38&-119&-129\\-21&1&-63&2&-45&-7\\-7&-1&-9&10&-12&-33\\14&-13&173&-42&148&190\end{bmatrix}$

    [S]-307

    (c) SageMathCell 54

    $\begin{bmatrix}-413&-649&26&-7&-73&-513\\42&66&0&-6&8&48\\-77&-121&4&1&-14&-93\\63&99&-6&6&11&87\\-147&-231&6&6&-27&-178\\308&484&-18&0&57&356\end{bmatrix}$

(d) SageMathCell 55

$$\begin{bmatrix} 126 & -64 & -94 & 98 & -40 & 96 \\ -55 & -28 & 100 & 10 & -15 & 84 \\ 128 & -80 & -85 & 141 & -116 & -32 \\ -132 & 112 & -13 & 3 & -70 & 150 \\ -117 & 27 & 97 & 106 & -110 & 81 \\ -149 & -54 & -66 & 132 & 46 & 111 \end{bmatrix}$$

[$\mathbb{A}$]-353

Answers

unique linear combination For an arbitrary vector $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ in $\mathbb{R}^n$,

$$\begin{aligned} \mathbf{b} &= \begin{bmatrix} b_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ b_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ b_3 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ b_n \end{bmatrix} \\ &= b_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + b_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + b_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + b_n \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \\ &= b_1 I_{:,1} + b_2 I_{:,2} + \cdots + b_n I_{:,n} \end{aligned}$$

and no other linear combination of the columns of I equals $\mathbf{b}$ (because each coefficient of the linear combination affects one and only one entry of $\mathbf{b}$).

unique linear combinations of polynomials An arbitrary element of $\mathbb{P}_2(\mathbb{R})$ takes the form $p(t) = at^2 + bt + c$. To write p as a linear combination of the elements of $T = \{t, t^2, 1 + t^2\}$, we need real numbers α, β, γ such that

$$\alpha t + \beta t^2 + \gamma(1 + t^2) = at^2 + bt + c$$

or $\gamma + \alpha t + (\gamma + \beta)t^2 = at^2 + bt + c$. From here, it is clear that we need $\gamma = c$, $\alpha = b$ and $\gamma + \beta = a$. Solving this last equation for β, we find $\beta = a - \gamma = a - c$. The algebra shows that not only is this a solution to the problem, it is the only one!

4.3 Functions and Transformations

Background

Give yourself a moment to recall everything you can about functions before reading on. Go ahead. Think about it. Close your eyes. Close the book and just think. There are no right or wrong answers. You remember what you remember.

What did you come up with? Common answers include things like "inputs and outputs", "black box", "slope-intercept", "graphing", "zeros", "$f(x)$", "there's a domain", and so on. Whatever came to mind is okay—it was probably related. Now try to divorce yourself from all those ideas. Free your mind from past experiences, and start over. We'll come back to the familiar ideas of function notation and graphing later. For now try to think more generally, more abstractly about functions.

A function is made from three ingredients—three sets, really. Two of the sets may contain any types of objects—real numbers in each, fruits in one and colors in the other, car companies in one and countries in the other, subsets of real numbers in one and polynomials in the other, matrices in one and integers in the other—no restrictions. The definition of function does not specify. These two sets are called the domain and codomain. Any set can be the domain of a function, and any set can be the codomain of a function. The only requirements of a function are placed on the third set. This set must contain exactly one ordered pair for each element of the domain. The order of the elements is important too. The first component of each ordered pair must be an element of the domain and the second component must be an element of the codomain.

A relation, like a function is made from three sets, a domain, a codomain, and a set of ordered pairs where the first component of the orderd pair is an element of the domain and the second component is an element of the codomain. Unlike a function, there are no further requirements. Thus, a relation can be thought of as a **rela**xed func**tion**. It has to adhere to fewer rules. Remember, the set of ordered pairs defining a function must contain exactly one ordered pair for each element of the domain, but a relation does not have this restriction.

To be precise, even though a relation is composed of three sets, A, B, C, where C is any subset of $\{(a,b) : a \text{ is in } A \text{ and } b \text{ is in } B\}$, the set C is the **relation** itself. The sets A and B are just ingredients in the definition of C. To simplify the notation, the set $\{(a,b) : a \text{ is in } A \text{ and } b \text{ is in } B\}$ is denoted $A \times B$, read "A cross B". To rephrase then, if A and B are sets, a relation is any subset of $A \times B$, and a **function** is a subset of $A \times B$ containing exactly one ordered pair for each element of A.

A relation C might be given using terse set notation $C = \{(a,b) \in A \times B : \text{rule of correspondence}\}$ such as in

$$C = \{(a,b) \in \mathbb{R} \times \mathbb{R} : a^2 + b^2 = 1\}.$$

The same relation might be simplified to just $a^2 + b^2 = 1$ if it is understood without writing explicitly that a and b are real numbers.

When a relation is a function, we may emphasize this point using the notation $f : A \to B$, read "f is a function from A to B". Implied in this notation is that f is a function and there exists a rule of correspondence specifying exactly which ordered pairs $(a,b) \in A \times B$ are in f. To define a specific function, the familiar function notation is often used, as in $f : A \to B$, $f(a) =$ "insert formula here", meaning $f = \{(a,b) \in A \times B : b = \text{"insert formula here"}\}$.

The domain and codomain of a function are often taken for granted, fading into the background in favor of focusing on the rule of correspondence, such as in $f(x) = 3x + 11$. It is just assumed or implied that the codomain is the set of all real numbers and the domain is some subset of the real numbers—with good reason. It would be rather repetitive to write $f : \mathbb{R} \to \mathbb{R}$ every time a function on the real numbers came up. More importantly, though, the domains of many functions with succinct formulas do not contain all real numbers. The domain only includes numbers that correspond to some element of the codomain. It would be a difficult distraction to have to write $f : A \to \mathbb{R}$ and get the set A correct every time a function was mentioned. To be complete, though, the definition of a function should include this information, and it is a convenience not to require it. To compensate for this lack of completeness in practice, a classic problem in algebra is to find the *implied* domain of functions such as $f(x) = \frac{3x+7}{2x-5}$—the subset of real number inputs that correspond to real number outputs according to the formula. Can you find the implied domain of $f(x) = \frac{3x+7}{2x-5}$? Answer on page 135.

The discussion of inverse functions, which will be very important for us, is made abstruse by not introducing the notion of a relation. The inverse of a relation is a relatively simple matter. If C is a relation with domain A and codomain B, the relation with domain B and codomain A given by $C^{-1} = \{(b,a) : (a,b) \text{ is in } C\}$ is called the **inverse relation** of C. Every relation has an inverse relation, no exceptions. Simply reversing the order of the ordered pairs in any relation gives its inverse. Consequently every function (which is just a special type of relation) has an inverse *relation*, and the inverse is conceptually straightforward. That is not to say that the inverse relation is always a function, however. There are plenty of functions that do not have inverse functions, and tackling this problem before tackling the simpler one of inverse relations is the root of much confusion and difficulty.

Maps and Transformations

Map, **mapping**, and **transformation** are all synonyms for function. All four words share the same definition. When the codomain is not the set of real numbers, the word "function" is often supplanted by one of the others. Getting used to this fact is a matter of experience.

The $3n + 1$ conjecture is that iteration (computing the sequence $n, R(n), R(R(n)), R(R(R(n))), \ldots$) of the *mapping* $R : \mathbb{Z}^+ \to \mathbb{Z}^+$ defined by

$$R(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ 3n+1 & \text{if } n \text{ is odd} \end{cases}$$

always ends with the cycle $4, 2, 1, 4, 2, 1, \ldots$. Try it starting with $n = 13$, for example. How many iterations does it take to get the first 4? Answer on page 135. Don't get too distracted, though. The point here is that R is a mapping.

The *transformation* $\mathcal{A} : C(D) \to C^1(D)$ given by $\mathcal{A}(f) = \int_0^x f(t)\,dt$ is a standard of calculus, though it is less common to discuss the antiderivative as a transformation during a calculus class. The derivative provides another example of a *transformation* from some set of functions to another set of functions. For example, $D : \mathbb{P}_n(\mathbb{R}) \to \mathbb{P}_{n-1}(\mathbb{R})$ defined by $D(p) = p'(x)$ is a *map* from the set of polynomials of degree at most n to the set of polynomials of degree at most $n - 1$. For example, if $p(x) = 3x^2 - 2x + 1$, then $D(p) = p'(x) = 6x - 2$. Mechanically there is no advantage to writing the derivative as a *transformation*, but conceptually it gives a certain perspective on the process of differentiation. The process itself can be thought of as a function!

The *map* $l : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ given by $l(s_0, s_1, s_2, \ldots) = s_1, s_2, s_3, \ldots$ is sometimes called the (left) shift operator. Its cousin, the left bit shift operator (called left-shift) plays a huge role in computing.

The determinant is a function or *map* $\det : \mathcal{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ for each n since each $n \times n$ matrix has exactly one determinant. The *transformation* $Z : GL_n(\mathbb{F}) \to GL_n(\mathbb{F})$ *mapping* an invertible matrix to its inverse provides motivation to think of finding the inverse of a matrix as a function as well. The most common *transformation* considered in linear algebra, however, is the *transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$, $T(\mathbf{x}) = M\mathbf{x}$ for some matrix M. Transformations T defined this way have two properties:

1. $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ for any $\mathbf{x}$ and any $\mathbf{y}$ in $\mathbb{R}^n$.

2. $T(c\mathbf{x}) = cT(\mathbf{x})$ for any c in $\mathbb{R}$ and $\mathbf{x}$ in $\mathbb{R}^n$.

Can you justify this claim? Answer on page 135. These two properties mean that performing addition and scalar multiplication in the domain and then transforming (the lefthand sides of the equations) gives the same result as transforming first and then performing the addition and scalar multiplication in the codomain (the righthand sides of the equations). For this reason we say this type of transformation preserves the operations of addition and scalar multiplication. Its properties form the essence for the abstraction of this idea to arbitrary vector spaces.

Linear Transformations

Given vector spaces V and W, a **linear transformation** is any transformation $L : V \to W$ such that for every x, y in V and scalar c,

1. $L(x + y) = L(x) + L(y)$ and

2. $L(cx) = cL(x)$.

This definition is modeled after $T : \mathbb{R}^n \to \mathbb{R}^m$, $T(\mathbf{x}) = M\mathbf{x}$, making T the canonical example of a linear transformation. Some of the other transformations mentioned in this section are linear transformations and some are not. For example, R is not since $R(2+3) = 16$ but $R(2)+R(3) = 1+10 = 11$, and $11 \neq 16$. It is easy to find positive integers that violate property 1. The derivative is a linear transformation since two basic results of calculus are that $(f + g)' = f' + g'$ and $(cf)' = cf'$. These rules of differentiation say precisely that properties 1 and 2 hold when differentiation is viewed as a mapping. In other words, $D(f + g) = D(f) + D(g)$ and $D(cf) = cD(f)$. What about $\mathcal{A}, l, \det$, and Z? Are they linear? Answers on page 135.

Vocabulary of Transformations

If $T : A \to B$ is a transformation, then

- $T(a)$ is called **the image** of a.
- If S is a subset of A, the set of all images, $\{T(x) : x \text{ is in } S\}$ is called **the image** of S.
- The set of all images, $\{T(a) : a \text{ is in } A\}$, is called **the range** of T, denoted range(T).
- a is called **a preimage** of b whenever $T(a) = b$.
- The set of all preimages of an element b of B is also called **the preimage** of b.
- the **kernel** of T is the preimage of $\mathbf{0}$.
- When the inverse relation of T is a function, it is denoted $T^{-1} : \text{range}(T) \to A$ and T^{-1} is called the **inverse function** of T or simply **the inverse** of T.

It is not an accident that the notation T^{-1} is used for the inverse of T. The notion of an inverse function extends the idea of multiplicative (and additive) inverses. By definition, if T has an inverse function T^{-1}, then $T^{-1}(T(a)) = a$ for any a in A. Can you justify this claim? Answer on page 136. Start with a, end with a. Composing T^{-1} with T returns the starting value, much like $M^{-1}(MA) = A$—left-multiplying the matrix A by invertible matrix M and then left-multiplying the result by the matrix M^{-1} results in the starting value A. Start with A, end with A.

Key Concepts

Cartesian product given any sets A and B, the Cartesian product of A and B is the set $\{(a, b) : a \in A \text{ and } b \in B\}$, denoted $A \times B$.

relation given any sets A and B, a relation is a subset C of $A \times B$.

function given any sets A and B, a function is a relation C such that if $(a, b_1) \in C$ and $(a, b_2) \in C$ then $b_1 = b_2$.

domain the set A in the definitions of relation and function.

codomain the set B in the definitions of relation and function.

rule of correspondence the rule that defines the set C in the definitions of relation and function.

inverse relation (of a relation T with domain A and codomain B) the relation with domain B, codomain A, and rule of correspondence $\{(T(a), a) : a \text{ is in } A\}$.

inverse function the inverse relation of a function whenever it happens to be a function.

inverse an inverse function. If $T^{-1} : \text{range}(T) \to A$ is the inverse of $T : A \to B$, then $T^{-1}(T(a)) = a$ for all a in A.

invertible a function is called invertible if its inverse relation is a function.

kernel the preimage of $\mathbf{0}$ whenever the codomain of a transformation is a vector space.

map a function, usually used when the codomain is not $\mathbb{R}$.

mapping a function, usually used when the codomain is not $\mathbb{R}$.

transformation a function, usually used when the codomain is not $\mathbb{R}$.

image an element, $T(a)$, of the codomain of a transformation T. Also, if S is a subset of the domain of T, the set $\{T(x) : x \text{ is in } S\}$.

preimage if b is an element of the codomain of a transformation T, then an element a in the domain of T is a preimage of b whenever $T(a) = b$. Also, the set of all elements a in the domain of T such that $T(a) = b$.

range the set of all images, $\text{range}(f) = \{f(a) : a \text{ is in } A\}$ for any function $f : A \to B$.

linear transformation given vector spaces V and W, a transformation $L : V \to W$ is linear (is a linear transformation) if for any x, y in V and any scalar c,

1. $L(x + y) = L(x) + L(y)$ and
2. $L(cx) = cL(x)$.

Exercises

1. Let $A = \{a, b, c, d\}$ and $B = \left\{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}\right\}$. Is the set a relation from A to B?
   (a) $\left\{\left(a, \frac{1}{2}\right), \left(b, \frac{1}{3}\right), \left(c, \frac{1}{4}\right), \left(d, \frac{1}{5}\right)\right\}$
   (b) $\left\{\left(a, \frac{1}{2}\right), \left(b, \frac{1}{3}\right), \left(c, \frac{1}{4}\right)\right\}$ [𝔸]-353
   (c) $\left\{\left(a, \frac{1}{2}\right), \left(a, \frac{1}{3}\right), \left(a, \frac{1}{4}\right), \left(a, \frac{1}{5}\right)\right\}$
   (d) $\left\{\left(a, \frac{1}{2}\right), (b, b), \left(\frac{1}{5}, c\right)\right\}$ [𝔸]-353
   (e) $\left\{\left(\frac{1}{2}, a\right), \left(\frac{1}{3}, b\right), \left(c, \frac{1}{3}\right), \left(d, \frac{1}{3}\right)\right\}$
   (f) $\left\{\left(a, \frac{1}{3}\right), \left(b, \frac{1}{3}\right), \left(c, \frac{1}{3}\right), \left(d, \frac{1}{3}\right)\right\}$ [𝔸]-353
2. In question 1, is the set a function from A to B? [𝔸]-353
3. The set $\left\{\left(\begin{bmatrix} 2 \\ 3 \end{bmatrix}, -7\right), \left(\begin{bmatrix} -\frac{1}{3} \\ 8 \end{bmatrix}, \frac{5}{7}\right), \left(\begin{bmatrix} 2\pi \\ 3 \end{bmatrix}, \sqrt{3}\right), \left(\begin{bmatrix} -2 \\ 221 \end{bmatrix}, -50\right)\right\}$ is a relation.
   (a) State a possible domain.
   (b) State a possible codomain.
   (c) State the range.
4. The set $\left\{(1, t), \left(2, t^2\right), \left(3, t^3\right), \left(4, t^4\right), \left(5, t^5\right)\right\}$ is a relation. [𝕊]-308
   (a) State a possible domain.
   (b) State a possible codomain.
   (c) State the range.
5. Redo question 3 where the set is a function instead of a relation.
6. Redo question 4 where the set is a function instead of a relation. [𝕊]-308
7. Let p be the mapping $p : \mathbb{Z}^+ \to \mathbb{Q}$, $p(z) = \frac{1}{z+z^2}$.
   (a) Find the image of 3.
   (b) Find the image of 23.
   (c) Find a preimage of $\frac{1}{6}$.
   (d) Find the preimage of $\frac{1}{6}$.
   (e) Find the image of $\{1, 2, 3\}$.
8. Let p be the mapping $p : \mathbb{Z} \to \mathbb{Q}$, $p(z) = \frac{z}{1+z^2}$. [𝕊]-308
   (a) Find the image of 3.
   (b) Find the image of 23.
   (c) Find a preimage of $\frac{2}{5}$.
   (d) Find the preimage of $\frac{2}{5}$.
   (e) Find the image of $\{1, 2, 3\}$.
9. Relation $R = \left\{\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \end{bmatrix}\right), \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -5 \end{bmatrix}\right)\right\}$ is a subset of $\mathbb{R}^2 \times \mathbb{R}^2$. Find R^{-1}.
10. Relation $R = \left\{\left(t^2 + 3, \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}\right), \left(3t^2 - 2t, \begin{bmatrix} 3 \\ -2 \\ 0 \end{bmatrix}\right)\right\}$ is a subset of $\mathbb{P}_2(\mathbb{R}) \times \mathbb{R}^3$. Find R^{-1}. [𝕊]-308
11. For the transformation T it is known that $T\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix}$, $T\left(\begin{bmatrix} -1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix}$, $T\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$, and that T is invertible.
    (a) Find a preimage of $\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. [𝔸]-353
    (b) Find $T^{-1}\left(\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}\right)$. [𝔸]-353
    (c) Find a preimage of $\begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix}$.
    (d) Find $T^{-1}\left(\begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix}\right)$.
    (e) Make a general statement about preimages and inverses.
12. Find the image of $\mathbf{a}$ under the mapping $T : \mathbb{R}^n \to \mathbb{R}^m$, $T(\mathbf{v}) = M\mathbf{v}$.
    (a) $M = \begin{bmatrix} 8 & -10 & -6 \\ 10 & 5 & 3 \end{bmatrix}$; $\mathbf{a} = \begin{bmatrix} -4 \\ 1 \\ 5 \end{bmatrix}$
    (b) $M = \begin{bmatrix} -9 & 1 \\ 3 & 0 \end{bmatrix}$; $\mathbf{a} = \begin{bmatrix} -7 \\ 10 \end{bmatrix}$ [𝕊]-308
    (c) $M = \begin{bmatrix} 7 & -3 \\ -9 & -1 \\ 8 & -4 \end{bmatrix}$; $\mathbf{a} = \begin{bmatrix} -3 \\ -6 \end{bmatrix}$
    (d) $M = \begin{bmatrix} 2 & 9 & 8 \\ -6 & 0 & -1 \\ -10 & -9 & 10 \end{bmatrix}$; $\mathbf{a} = \begin{bmatrix} 10 \\ 6 \\ 5 \end{bmatrix}$
13. Show that $f : \mathbb{R} \to \mathbb{R}$ is not a linear transformation.
    (a) $f(x) = e^x$
    (b) $f(x) = \ln x$ [𝔸]-353
    (c) $f(x) = x^2$
    (d) $f(x) = \sqrt{x - 3}$
14. Show that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation.
    (a) $T(x) = 13x$
    (b) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1 & 3 \\ -2 & 5 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$
    (c) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 0 & 1 \\ -1 & -4 \\ 4 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$
    (d) $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 4 & -1 & -5 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ [𝔸]-353
15. Show that $L : \mathbb{P}_m(\mathbb{R}) \to \mathbb{P}_n(\mathbb{R})$ is a linear transformation.
    (a) $L(ax + b) = \frac{1}{2}ax^2 + bx + c$
    (b) $L(ax^2 + bx + c) = 6ax + 3b$ [𝔸]-353
16. Perhaps ironically, what is known as a linear function in algebra is not necessarily a linear transformation. Verify that $f : \mathbb{R} \to \mathbb{R}$, $f(x) = mx + b$
    (a) is a linear transformation when $b = 0$

(b) is not a linear transformation when $b \neq 0$

17. Show that $\det : \mathcal{M}_{2\times 2}(\mathbb{R}) \to \mathbb{R}$ is not a linear transformation.

18. $T : \mathbb{R}^2 \to \mathbb{R}^3$ is linear, $T\left(\begin{bmatrix} -1 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} -2 \\ 5 \\ 4 \end{bmatrix}$, and $T\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} -3 \\ -5 \\ 1 \end{bmatrix}$. Find

    (a) $T\left(\begin{bmatrix} -2 \\ 8 \end{bmatrix}\right)$

    (b) $T\left(\begin{bmatrix} 6 \\ -3 \end{bmatrix}\right)$ [A]-353

    (c) $T\left(\begin{bmatrix} 1 \\ 3 \end{bmatrix}\right)$

    (d) $T\left(\begin{bmatrix} -3 \\ 5 \end{bmatrix}\right)$ [A]-353

19. $T : \mathbb{R}^3 \to \mathbb{R}^2$ is linear, $T\left(\begin{bmatrix} -2 \\ 0 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} -3 \\ -5 \end{bmatrix}$, and $T\left(\begin{bmatrix} 5 \\ -4 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} -5 \\ 1 \end{bmatrix}$. Find

    (a) $T\left(\begin{bmatrix} 4 \\ 0 \\ -8 \end{bmatrix}\right)$ [A]-353

    (b) $T\left(\begin{bmatrix} 15 \\ -12 \\ -3 \end{bmatrix}\right)$

    (c) $T\left(\begin{bmatrix} 1 \\ -4 \\ 7 \end{bmatrix}\right)$ [A]-354

    (d) $T\left(\begin{bmatrix} -7 \\ -4 \\ 5 \end{bmatrix}\right)$

20. Can you find a matrix M such that the transformation T of question 18 can be expressed as $T(\mathbf{x}) = M\mathbf{x}$? [A]-354

21. Can you find a matrix M such that the transformation T of question 19 can be expressed as $T(\mathbf{x}) = M\mathbf{x}$?

22. Let $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 0 & -2 & 6 \\ -7 & -4 & 5 \\ 1 & 2 & -5 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}$. Find the preimage of $\begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$. Is T invertible? Explain.

23. The range of the function $g : \mathbb{R} \to \mathbb{R}$, $g(x) = x^2$ is $[0, \infty)$. Therefore, the inverse relation of g is

$$g^{-1} = \{(y, x) \in [0, \infty) \times \mathbb{R} : y = x^2\}.$$

    Verify that g is a function and argue that g^{-1} is not.

24. Justify the claim.

    (a) For any function $C \subseteq A \times B$, if $(a, b_1) \in C$ and $(a, b_2) \in C$ then $b_1 = b_2$.

    (b) If the mapping $L : V \to W$ is linear, then $L(\mathbf{0}) = \mathbf{0}$. Note: the $\mathbf{0}$ on the lefthand side is the zero vector of V and the $\mathbf{0}$ on the righthand side is the zero vector in W. They are not necessarily equal.

    (c) The composition of linear transformations is linear.

    (d) A map $L : V \to W$ is linear if and only if $L(x + cy) = L(x) + cL(y)$ for every x, y in V and scalar c.

Answers

implied domain The implied domain is the set of all inputs that correspond to real number outputs. The function $f(x) = \frac{3x+7}{2x-5}$ includes a fraction, a quantity that is undefined precisely when the denominator is zero. Hence we must discard all numbers that satisfy $2x - 5 = 0$. Solving this equation for x is a simple matter: $x = \frac{5}{2}$ and so the implied domain is $\mathbb{R}\backslash\left\{\frac{5}{2}\right\}$ (all real numbers except $\frac{5}{2}$).

$3n + 1$ problem $R(13) = 3(13)+1 = 40$ since 13 is odd. $R(R(13)) = R(40) = \frac{40}{2} = 20$ since 40 is even. $R(R(R(13))) = R(R(40)) = R(20) = 10$, and so on. The sequence $n, R(n), R(R(n)), R(R(R(n))), \ldots$ is

$$13, 40, 20, 10, 5, 16, 8, 4, 2, 1, 4, 2, 1, 4, 2, \ldots$$

so it takes 7 iterations to get the first 4.

two properties For $T : \mathbb{R}^n \to \mathbb{R}^m$, $T(\mathbf{x}) = M\mathbf{x}$ where M is a matrix,

1. $T(\mathbf{x} + \mathbf{y}) = M(\mathbf{x} + \mathbf{y}) = M\mathbf{x} + M\mathbf{y} = T(\mathbf{x}) + T(\mathbf{y})$; and
2. $T(c\mathbf{x}) = M(c\mathbf{x}) = c(M\mathbf{x}) = cT(\mathbf{x})$.

linear or not? Another two results of calculus are that $\int(f+g) = \int f + \int g$ and $\int(cf) = c\int f$, so $\mathcal{A}(f+g) = \mathcal{A}(f) + \mathcal{A}(g)$ and $\mathcal{A}(cf) = c\mathcal{A}(f)$. $\mathcal{A}$ is a linear transformation. On one hand,$l(s+t) = l\left((s_0, s_1, s_2, \ldots) + (t_0, t_1, t_2, \ldots)\right) =$

$l(s_0 + t_0, s_1 + t_1, s_2 + t_2, \ldots) = s_1 + t_1, s_2 + t_2, \ldots$. On the other hand, $l(s) + l(t) = l(s_0, s_1, s_2, \ldots) + l(t_0, t_1, t_2, \ldots) = (s_1, s_2, \ldots) + (t_1, t_2, \ldots) = s_1 + t_1, s_2 + t_2, \ldots$. Therefore $l(s + t) = l(s) + l(t)$. Property 1 holds for l. To check that property 2 holds, note that $l(cs) = l(cs_0, cs_1, cs_2, \ldots) = cs_1, cs_2, \ldots$ and $cl(s) = cl(s_0, s_1, s_2, \ldots) = c(s_1, s_2, \ldots) = cs_1, cs_2, \ldots$ too. det is not a linear transformation. For example, let $A = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$ and $B = -A$. Then $\det A = \det B = 1$, so $\det(A) + \det(B) = 2$ while $\det(A + B) = 0$. Matrix inversion is also not a linear transformation. Using the same matrices A and B, $A^{-1} = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}$ and $B^{-1} = \begin{bmatrix} -5 & 3 \\ 3 & -2 \end{bmatrix}$ so the sum of the inverses is $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ but the inverse of the sum does not exist. In other words, $V(A) + V(B) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ but $V(A + B)$ does not exist, so they are not equal.

definition of inverse Since a is in the domain of T, there must be an ordered pair (a, b) in T—meaning $T(a) = b$. By definition (b, a) is then in the inverse of T—meaning $T^{-1}(b) = a$. Putting these facts together, $T^{-1}(T(a)) = T^{-1}(b) = a$.

4.4 Linear Transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$

Geometric Interpretation of Linear Transformations

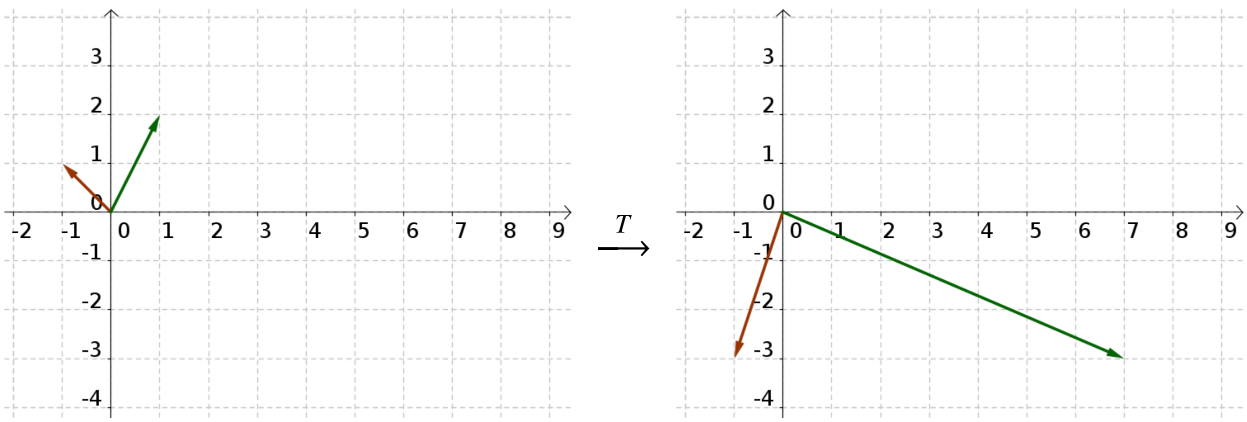
Drawing vectors as arrows, as in section 1.4, gives us a way to picture linear transformations. We can draw any vector and its image to help understand the action of the map geometrically. For example, if every vector we draw has an image pointing in the same direction but twice as long, that gives us a clear picture of how it transforms vectors. It doubles their magnitudes.

Since vectors do not have an inherent starting location, we can always imagine them starting anywhere. In the case of picturing linear transformations, it is helpful to imagine vectors rooted at the origin. Much like plotting points in the plane, these vectors are marked off starting at $(0,0)$. Given this special use of the vector, there is little distinction between the point at the head of an arrow and the arrow itself. For this reason, it is just as common to imagine a vector as a point in the plane as it is to imagine it as an arrow.

Consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$,

$$T(\mathbf{v}) = \begin{bmatrix} 3 & 2 \\ 1 & -2 \end{bmatrix} \mathbf{v}.$$

The image of $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ is $\begin{bmatrix} 3 & 2 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -3 \end{bmatrix}$ and the image of $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is $\begin{bmatrix} 3 & 2 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 7 \\ -3 \end{bmatrix}$. Geometrically these facts are captured by the diagram



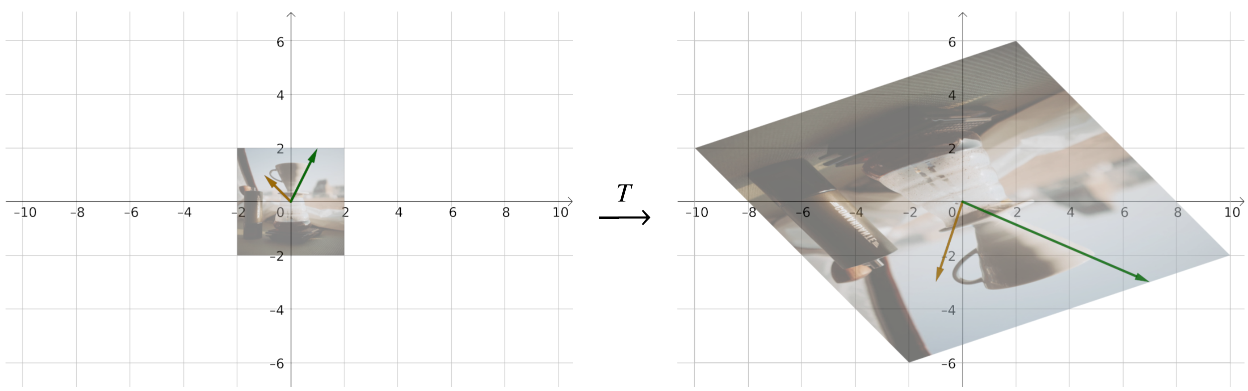
where the vectors have been represented by arrows rooted at the origin and the T indicates that the change is due to the transformation T. The vectors have also been color coded so the image of the brown vector is brown and the image of the green vector is green. From just these two sample vectors, it is hard to describe just what the transformation does in general. This is where it is helpful to interpret vectors as points. If we color a bunch of points, transform each of them, one at a time, giving their images the same color, we get a much clearer picture of the action of the transformation.

Consider again the transformation

$$T(\mathbf{v}) = \begin{bmatrix} 3 & 2 \\ 1 & -2 \end{bmatrix} \mathbf{v}$$

but this time imagine the vectors it acts on and their images as points. Coloring all the points in the square with opposite corners at $(-2,-2)$ and $(2,2)$ to manifest as a photo of a coffee mug[3] and coloring their images accordingly is summarized in the folowing diagram.

[3] Photo by Dziana Hasanbekava from Pexels



The same brown and green vectors as before are superimposed on the picture to help relate back to this interpretation. The point at the end of the green arrow is sky blue and lands on the boundary of the picture before transforming, so the point at the end of the green arrow is sky blue and lands on the boundary of the picture after transforming, too. A large portion of the green vector runs up the side of the coffee mug both before and after transformation. The point at the end of the brown arrow is sky blue and lands just next to the handle of the coffee mug before transforming, so the point at the end of the brown arrow is sky blue and lands just next to the handle of the coffee mug after transforming, too. In all, the transformed image is larger than the original, rotated, reflected (the handle is on the left of the coffee mug in one picture and on the right in the other), and sheared (the transformed picture covers a parallelogram, not a square). These are the words we use to describe the action of the transformation. It scales, rotates, reflects, and shears the plane, and objects in it. Actually, these are the only invertible actions a linear transformation can take on the plane, as we will see.

The Matrix of a Linear Transformation

Suppose a map $G : \mathbb{R}^n \times \mathbb{R}^m$ is given by $G(\mathbf{v}) = M\mathbf{v}$ for some matrix M. Then, by theorem 2 part 5, $G(\mathbf{u} + \mathbf{v}) = M(\mathbf{u} + \mathbf{v}) = M\mathbf{u} + M\mathbf{v} = G(\mathbf{u}) + G(\mathbf{v})$ and by theorem 3 part 4, $G(c\mathbf{u}) = M(c\mathbf{u}) = c(M\mathbf{u}) = cG(\mathbf{u})$ for any vectors $\mathbf{u}$ and $\mathbf{v}$ and any scalar c. Therefore G is a linear transformation.

Suppose a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by matrix multiplication by an $m \times n$ matrix A. That is, $T(\mathbf{v}) = A\mathbf{v}$. Then it is easy to calculate its action on the vectors

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

the columns of the $n \times n$ identity matrix. Thinking of matrix multiplication of a vector as a linear combination of the columns of the matrix, it is clear that

$$A\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = A_{:,1},\ A\begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = A_{:,2},\ A\begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} = A_{:,3}, \ldots, A\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = A_{:,n}.$$

In short, $AI_{:,j} = A_{:,j}$. The j^{th} column of A is the image of the j^{th} column of I.

On the other hand, suppose you know the images of the columns of I but are not given T as a matrix product. All you know is

$$T(I_{:,1}) = \mathbf{c}_1,\ T(I_{:,2}) = \mathbf{c}_2, \ldots, T(I_{:,n}) = \mathbf{c}_n$$

for some $m \times 1$ vectors $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$. It turns out this is enough information to determine T, and T can be represented by matrix multiplication! Let $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ be an arbitrary $n \times 1$ vector. Due to linearity,

$$\begin{aligned}
T(\mathbf{v}) &= T\left(\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}\right) \\
&= T\left(\begin{bmatrix} v_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ v_2 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ v_n \end{bmatrix}\right)
\end{aligned}$$

$$\begin{aligned}
&= T\left(v_1\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + v_2\begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + v_n\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}\right) \\
&= T\left(v_1 I_{:,1} + v_2 I_{:,2} + \cdots + v_n I_{:,n}\right) \\
&= T\left(v_1 I_{:,1}\right) + T\left(v_2 I_{:,2}\right) + \cdots + T\left(v_n I_{:,n}\right) \\
&= v_1 T\left(I_{:,1}\right) + v_2 T\left(I_{:,2}\right) + \cdots + v_n T\left(I_{:,n}\right) \\
&= v_1 \mathbf{c}_1 + v_2 \mathbf{c}_2 + \cdots + v_n \mathbf{c}_n \\
&= \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix} \mathbf{v}.
\end{aligned}$$

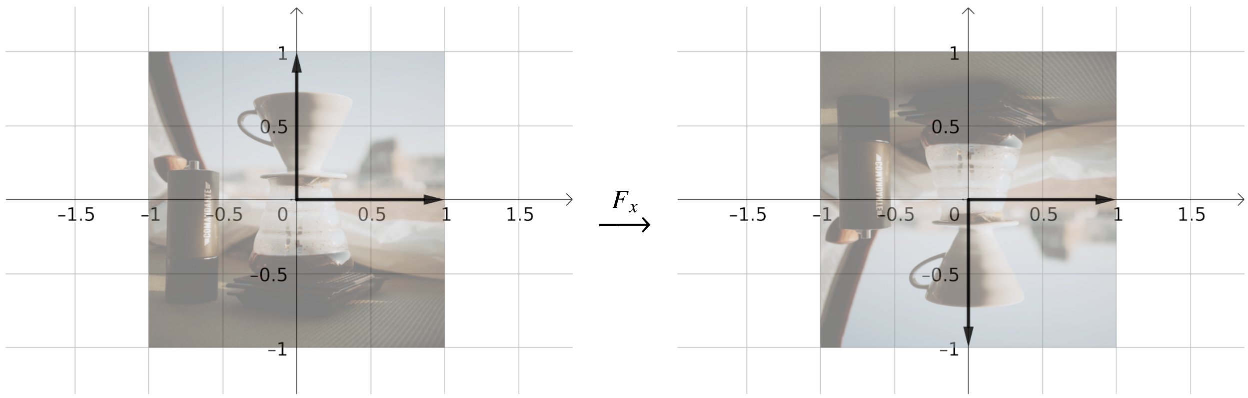
In other words, to represent T as a matrix product, form the matrix with columns equal to the images of the columns of I.

These calculations justify the folowing theorem.

Theorem 10. [The Standard Matrix of a Linear Transformation] *Given a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, T is linear if and only if $T(\mathbf{v}) = M\mathbf{v}$ where $M_{:,j} = T(I_{:,j}),\ j = 1, 2, \ldots, n$. M is called the standard matrix of T.*

In words, a transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is linear if and only if it can be represented by multiplication by a matrix whose columns are the images of the columns of the identity matrix.

This observaton can be applied immediately to write down the algebraic (matrix) representation of transformations with which you may already be familiar. For example, consider reflection about the x-axis in the plane, call it F_x. Geometrically, this transformation can be illustrated as in the following diagram.



The image of $I_{:,1}$ is $I_{:,1}$ and the image of $I_{:,2}$ is $-I_{:,2}$, so *if* this is a linear transformation,

$$F_x(\mathbf{v}) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{v}.$$

It is easy to verify that F_x acts on the entire plane (not just the columns of I) in the way expected.

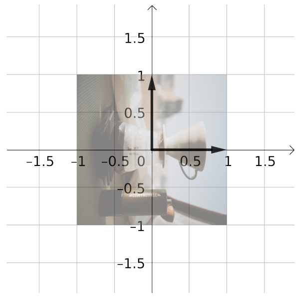
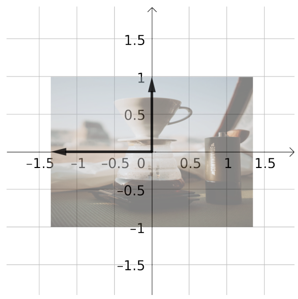
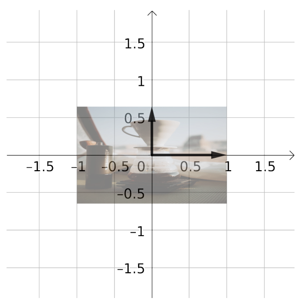
$$F_x(\mathbf{v}) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{v} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ -v_2 \end{bmatrix}$$

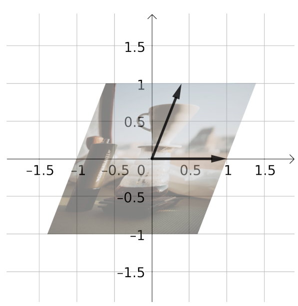
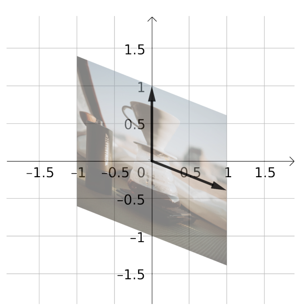
so the image of $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ is $\begin{bmatrix} v_1 \\ -v_2 \end{bmatrix}$, the reflection of $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ about the x-axis. It must be that reflection about the x-axis is a linear transformation. Can you justify this using the definition of linear transformation? Answer on page 146.

Notice that $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ is an elementary matrix (scale the second row by -1). Every elementary 2×2 matrix takes one of the following five forms—swap, scale first row, scale second row, replace first row, replace second row, respectively—for some scalar r or $s \neq 0$:

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

The following series of diagrams illustrates the types of transformations attainable by multiplication by these elementary matrices. Geometrically they are reflection, scaling, scaling with reflection, and shearing.

matrix M	image under $T(\mathbf{v}) = M\mathbf{v}$	Notes
$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$		• reflection about the line $y = x$ • $I_{:,1}$ and $I_{:,2}$ swap places
$\begin{bmatrix} s & 0 \\ 0 & 1 \end{bmatrix}$		• horizontal scale by factor s • includes reflection about the y-axis when $s < 0$ • expansion when $\lvert s\rvert > 1$; compression when $\lvert s\rvert < 1$
$\begin{bmatrix} 1 & 0 \\ 0 & s \end{bmatrix}$		• vertical scale by factor s • includes reflection about the x-axis when $s < 0$ • expansion when $\lvert s\rvert > 1$; compression when $\lvert s\rvert < 1$

matrix M	image under $T(\mathbf{v}) = M\mathbf{v}$	Notes
$\begin{bmatrix} 1 & r \\ 0 & 1 \end{bmatrix}$		• horizontal shear by factor r • $I_{:,1}$ is unaffected
$\begin{bmatrix} 1 & 0 \\ r & 1 \end{bmatrix}$		• vertical shear by factor r • $I_{:,2}$ is unaffected

As we have seen previously, any invertible matrix can be written as a product of elementary matrices. If there were some connection between matrix multiplication and linear transformations, we would be on our way to a comprehensive characterization of linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$. By theorem 2 part 4 $A(B\mathbf{v}) = (AB)\mathbf{v}$. In terms of linear transformations, the left side, $A(B\mathbf{v})$, represents applying the transformation whose associated matrix is B first and then applying the transformation whose associated matrix is A to the result. In other words, $A(B\mathbf{v})$ represents composing the two transformations whose associated matrices are B and A. The right side, $(AB)\mathbf{v}$, represents applying the transformation whose associated matrix is the product AB. It must be that matrix multiplication corresponds to function composition! To facilitate the following discussion, which puts these words into symbols and elaborates, for any matrix M we adopt the notation T_M for the linear transformation $T_M : \mathbb{R}^n \to \mathbb{R}^m$, $T_M(\mathbf{v}) = M\mathbf{v}$.

Letting T_A be an arbitrary linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and T_B an arbitrary linear transformation from $\mathbb{R}^p$ to $\mathbb{R}^n$, the following calculation encapsulates the idea that matrix multiplication corresponds to function composition.

$$\begin{aligned}(T_A \circ T_B)(\mathbf{v}) &= T_A(T_B(\mathbf{v}))\\ &= T_A(B\mathbf{v})\\ &= A(B\mathbf{v})\\ &= (AB)\mathbf{v}\\ &= T_{AB}(\mathbf{v}).\end{aligned}$$

This calculation has two important consequences. First, if M is invertible, then $T_M \circ T_{M^{-1}} = T_{M^{-1}} \circ T_M = T_I$. In other words, $(T_M \circ T_{M^{-1}})(\mathbf{v}) = (T_{M^{-1}} \circ T_M)(\mathbf{v}) = T_I(\mathbf{v}) = I\mathbf{v} = \mathbf{v}$, so T_M is invertible and $(T_M)^{-1} = T_{M^{-1}}$. Second, if M is invertible and we write M as the product of elementary matrices, $E_1E_2\cdots E_p$, then

$$\begin{aligned}T_M(\mathbf{v}) &= M\mathbf{v}\\ &= (E_1E_2\cdots E_p)\mathbf{v}\\ &= \left(T_{E_1} \circ T_{E_2} \circ \cdots \circ T_{E_p}\right)(\mathbf{v}),\end{aligned}$$

so T_M is a composition of linear transformations whose associated matrices are elementary matrices.

Finally, because the action of transformations defined by 2×2 elementary matrices include only reflection, scaling, and shearing, these actions and compositions of them are the only actions of invertible linear transformations on $\mathbb{R}^2$.

So what about T_M where M is noninvertible? Noninvertible matrices have linearly dependent columns and linearly dependent rows. In the case of 2×2 matrices that means the rows are multiples of one another or one of the rows contains two zeros. Likewise, either its columns are multiples of one another or one of the columns contains two zeros. In any case, it has the form

$$N = \begin{bmatrix} ka & \ell a \\ kb & \ell b \end{bmatrix}$$

for some scalars a, b, k, ℓ. Can you verify this form covers all four cases? Answer on page 146. If $k \neq 0$

$$N = \begin{bmatrix} ka & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ kb & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & \frac{\ell}{k} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} ka & \ell a \\ kb & \ell b \end{bmatrix}$$

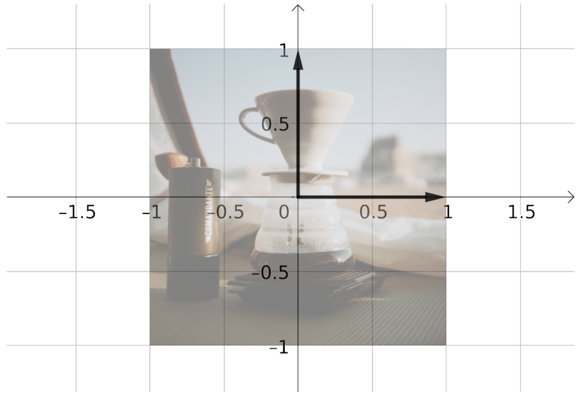
and if $k = 0$

$$N = \begin{bmatrix} 1 & 0 \\ 0 & \ell b \end{bmatrix}\begin{bmatrix} 1 & \ell a \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & \ell a \\ 0 & \ell b \end{bmatrix}.$$

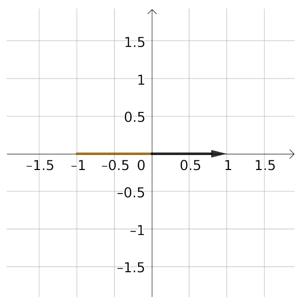
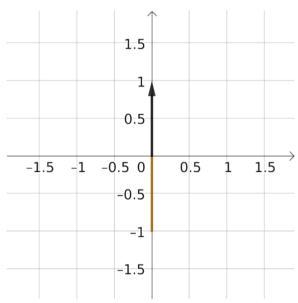
Either way, N can be written as the product of elementary matrices and either

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

These matrices are called **projection matrices**. Their action is to squash (or project) the entire plane onto the x-axis or the y-axis, respectively, as shown below acting on



as before. The brown line segment in the diagrams below indicates the part of the axis in the image that is not covered by the vector.

matrix M	image under $T(\mathbf{v}) = M\mathbf{v}$	Notes
$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$		• projection onto the x-axis • $I_{:,1}$ is unaffected • $I_{:,2}$ is squashed to the origin
$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$		• projection onto the y-axis • $I_{:,2}$ is unaffected • $I_{:,1}$ is squashed to the origin

These projections complete the characterization of linear transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$, also called linear operators on $\mathbb{R}^2$. Because every 2×2 matrix can be written as a product of the matrices

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} s & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & s \end{bmatrix}, \begin{bmatrix} 1 & r \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ r & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

and multiplication by these matrices represents reflection, scaling, shearing, and/or projection, we have the following theorem.

Theorem 11. [Characterization of Linear Transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$] *A transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ is linear if and only if it is a composition of some sequence of reflections, scalings, shearings, and projections.*

Key Concepts

geometric interpretation of vectors vectors in $\mathbb{R}^n$ are often thought of as points.

matrices and linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear if and only if $T(\mathbf{v}) = M\mathbf{v}$ where $M_{:j} = T(I_{:,j}),\ j = 1, 2, \ldots, n$.

elementary matrices as linear transformations of the plane the action of a swap matrix is reflection about the line $y = x$; the action of a scale matrix is scaling and possibly reflection; the action of a replacement matrix is shearing.

noninvertible linear transformations of the plane a linear transformation of the plane, T_M, is noninvertible if and only if M is the product of a projection matrix with elementary matrices.

characterization of linear transformations of the plane see theorem 11.

standard matrix see theorem 10.

projection matrix (onto an axis in $\mathbb{R}^2$) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ or $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.

Exercises

1. (i) Find the standard matrix of $T : \mathbb{R}^2 \to \mathbb{R}^2$; and (ii) find the image of $\begin{bmatrix} 1 & 1 \end{bmatrix}^T$.
   (a) T shears points so that the image of $(1, 0)$ is $\left(1, -\frac{1}{2}\right)$ while $(0, 1)$ is unaffected. [$\mathbb{A}$]-354
   (b) T reflects points through the origin.
   (c) T scales horizontally by a factor of 3. [$\mathbb{A}$]-354
   (d) T projects onto the y-axis.
2. Argue that even though the transformations in exercises 1b and 13c have different geometric descriptions, they are the same transformation.
3. Find the standard matrix of $T : \mathbb{R}^2 \to \mathbb{R}^2$.
   (a) T reflects points about the y-axis and then scales them both vertically and horizontally by a factor of 2. [$\mathbb{S}$]-308
   (b) T scales points vertically by a factor of $\frac{1}{2}$ and then shears them vertically by a factor of $\frac{1}{2}$.
   (c) T reflects points about the x-axis and then shears horizontally by a factor of 0.77.
   (d) T dilates points horizontally and vertically by a factor of $\frac{28}{9}$ and then shears horizontally by a factor of $\frac{3}{2}$. [$\mathbb{A}$]-354
   (e) T shears points vertically such that $\begin{bmatrix} 1 & 0 \end{bmatrix}^T$ maps to $\begin{bmatrix} 1 & 1.44 \end{bmatrix}$ and then reflects points about the origin.
   (f) T projects points onto the x-axis and then shears them vertically by a factor of 2. [$\mathbb{A}$]-354
   (g) T shears horizontally by a factor of 0.76 and then dilates points horizontally by a factor of $\frac{17}{6}$. [$\mathbb{A}$]-354
4. Are S and T the same transformation?
   (a) S rotates points (counterclockwise) about the origin by 45°, reflects them about the x-axis, and then rotates them clockwise about the origin by 45°.
   T reflects points about the line $y = -x$.
   (b) S reflects points about the y-axis and then rotates them $\frac{\pi}{4}$ radians clockwise about the origin.
   T rotates points $\frac{3\pi}{4}$ radians clockwise about the origin and then reflects them over the x-axis. [$\mathbb{S}$]-309
   (c) S shears points vertically by a factor of 2 and then scales them both horizontally and vertically by a factor of $\frac{1}{2}$.
   T scales points by horizontally and vertically by a factor of $\frac{1}{2}$ and then shears them vertically by a factor of 2.
   (d) S reflects points about the x-axis and then reflects them about the y-axis.
   T rotates points 180° about the origin. [$\mathbb{A}$]-354
5. Find the standard matrix for the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ such that
   (a) $T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -5.6 \\ 14.3 \end{bmatrix}$, $T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -2.7 \\ -7.4 \end{bmatrix}$, and $T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -11 \\ 4.5 \end{bmatrix}$.
   (b) $T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -7.5 \\ 1 \\ -13.2 \\ -3.7 \end{bmatrix}$ and $T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -5.2 \\ 10.1 \\ -4.3 \\ 4.3 \end{bmatrix}$. [$\mathbb{S}$]-309
   (c) $T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 17 \\ -94 \\ 50 \\ -30 \end{bmatrix}$, $T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -87 \\ 67 \\ 32 \\ 143 \end{bmatrix}$, and $T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -45 \\ -129 \\ -78 \\ -33 \end{bmatrix}$.
   (d) $T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -79 \\ 117 \\ 38 \end{bmatrix}$ and $T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 132 \\ 98 \\ -77 \end{bmatrix}$. [$\mathbb{A}$]-354

(e) $T\left(\begin{bmatrix}1\\0\\0\\0\end{bmatrix}\right) = \begin{bmatrix}-7\\7\end{bmatrix}$, $T\left(\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\right) = \begin{bmatrix}-14\\-8\end{bmatrix}$,
$T\left(\begin{bmatrix}0\\0\\1\\0\end{bmatrix}\right) = \begin{bmatrix}-4\\11\end{bmatrix}$, and $T\left(\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right) = \begin{bmatrix}-15\\-6\end{bmatrix}$.

6. Find a sequence of elementary and/or projection matrices whose product is the standard matrix of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$.

   (a) T reflects points about the y-axis and then projects them onto the x-axis.

   (b) T dilates points vertically by a factor of $\frac{11}{7}$ and then shears them horizontally by a factor of 1.62. [$\mathbb{A}$]-354

   (c) T projects points onto the y-axis and then dilates them horizontally and vertically by a factor of 2.

   (d) T reflects points about the x-axis and then shears them horizontally by a factor of 1.27. [$\mathbb{A}$]-354

   (e) T shears points horizontally such that $\begin{bmatrix}0 & 1\end{bmatrix}^T$ maps to $\begin{bmatrix}0.08 & 1\end{bmatrix}^T$ and then reflects them about the x-axis.

   (f) T contracts points horizontally and vertically by a factor of $\frac{2}{3}$ and then reflects them about the y-axis. [$\mathbb{A}$]-354

7. Though it is not strictly necessary, you may find SageMath helpful. Find the standard matrix for the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ such that

   (a) SageMathCell $T\left(\begin{bmatrix}41\\-3\\-2\end{bmatrix}\right) = \begin{bmatrix}7\\-4\end{bmatrix}$, $T\left(\begin{bmatrix}-6\\1\\0\end{bmatrix}\right) = \begin{bmatrix}-13\\-3\end{bmatrix}$, and $T\left(\begin{bmatrix}-5\\-1\\1\end{bmatrix}\right) = \begin{bmatrix}-9\\-13\end{bmatrix}$. [$\mathbb{S}$]-309

   (b) SageMathCell $T\left(\begin{bmatrix}1\\-3\end{bmatrix}\right) = \begin{bmatrix}-8\\5\\-3\\6\end{bmatrix}$ and $T\left(\begin{bmatrix}7\\-20\end{bmatrix}\right) = \begin{bmatrix}2\\-5\\3\\-4\end{bmatrix}$.

   (c) SageMathCell $T\left(\begin{bmatrix}-46\\17\\-95\end{bmatrix}\right) = \begin{bmatrix}-4\\-8\\-2\\8\end{bmatrix}$, $T\left(\begin{bmatrix}-4\\1\\-6\end{bmatrix}\right) = \begin{bmatrix}1\\0\\-6\\6\end{bmatrix}$, and $T\left(\begin{bmatrix}-5\\2\\-11\end{bmatrix}\right) = \begin{bmatrix}-7\\-1\\4\\2\end{bmatrix}$. [$\mathbb{A}$]-354

   (d) SageMathCell $T\left(\begin{bmatrix}1\\6\end{bmatrix}\right) = \begin{bmatrix}-2\\0\\8\end{bmatrix}$ and $T\left(\begin{bmatrix}3\\19\end{bmatrix}\right) = \begin{bmatrix}8\\4\\7\end{bmatrix}$.

   (e) SageMathCell $T\left(\begin{bmatrix}3\\-4\\9\\2\end{bmatrix}\right) = \begin{bmatrix}-9\\1\end{bmatrix}$, $T\left(\begin{bmatrix}-6\\14\\-18\\-5\end{bmatrix}\right) = \begin{bmatrix}5\\-5\end{bmatrix}$, $T\left(\begin{bmatrix}5\\-5\\16\\3\end{bmatrix}\right) = \begin{bmatrix}6\\-2\end{bmatrix}$, and $T\left(\begin{bmatrix}1\\-3\\3\\1\end{bmatrix}\right) = \begin{bmatrix}9\\2\end{bmatrix}$. [$\mathbb{A}$]-354

8. Argue that the transformation $S : \mathbb{R}^2 \to \mathbb{R}^2$, $S\left(\begin{bmatrix}x & y\end{bmatrix}^T\right) = \begin{bmatrix}x^2 & y^2\end{bmatrix}^T$ is not linear by the following steps.

   (a) Find the images of $I_{:,1}$ and $I_{:,2}$ under S.

   (b) Form the matrix M whose columns are the vectors from part 8a.

   (c) Find a vector $\mathbf{v}$ such that $S(\mathbf{v}) \neq M\mathbf{v}$.

9. Argue that the transformation $S : \mathbb{R}^2 \to \mathbb{R}^2$, $S\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \begin{bmatrix}2 & 3\\-4 & 1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} + \begin{bmatrix}-3\\2\end{bmatrix}$ is not linear by the following steps.

   (a) Find the images of $I_{:,1}$ and $I_{:,2}$ under S.

   (b) Form the matrix M whose columns are the vectors from part 9a.

   (c) Find a vector $\mathbf{v}$ such that $S(\mathbf{v}) \neq M\mathbf{v}$.

10. Show that reflection about the origin of the plane, which maps $\begin{bmatrix}x\\y\end{bmatrix}$ to $\begin{bmatrix}-x\\-y\end{bmatrix}$, is linear by finding a matrix M such that $M\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}-x\\-y\end{bmatrix}$. [$\mathbb{A}$]-354

11. Show that $F_{xy} : \mathbb{R}^2 \to \mathbb{R}^2$, $F_{xy}\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \begin{bmatrix}-x\\-y\end{bmatrix}$ (reflection about the origin of the plane) is linear by showing that $F_{xy}(\mathbf{u}+\mathbf{v}) = F_{xy}(\mathbf{u}) + F_{xy}(\mathbf{v})$ and $F_{xy}(c\mathbf{u}) = cF_{xy}(\mathbf{u})$ for any vectors $\mathbf{u}$ and $\mathbf{v}$ and scalar c.

Rotation. The remaining exercises explore rotation as a composition of elementary matrices and as a linear transformation itself. Each question after number 12 depends on the results of 12.

12. Argue that rotation in the plane is linear and write it algebraically as matrix multiplication by the following steps. Let R_θ denote rotation through angle θ counterclockwise about the origin. [𝕊]-310
   (a) Argue geometrically that R_θ is a linear transformation. Your argument need not be a proof, just enough reason to believe R_θ is (likely) linear.
   (b) Find the images of $I_{:,1}$ and $I_{:,2}$ under R_θ.
   (c) Write R_θ as a matrix product using the results of 12b.
   (d) Demonstrate algebraically that the matrix derived in part 12c affects rotation through angle θ about the origin on an arbitrary vector $\begin{bmatrix} x \\ y \end{bmatrix}$. This shows that rotation is properly represented by matrix multiplication, making it a linear transformation by theorem 10.
13. (i) Find the standard matrix of $T : \mathbb{R}^2 \to \mathbb{R}^2$; and (ii) find the image of $\begin{bmatrix} 1 & 1 \end{bmatrix}^T$.
   (a) T rotates points counterclockwise about the origin by $\frac{\pi}{2}$ radians. [𝕊]-312
   (b) T rotates points clockwise about the origin by $\frac{\pi}{4}$ radians.
   (c) T rotates points counterclockwise about the origin by 180 degrees. [𝔸]-354
14. Find the standard matrix of $T : \mathbb{R}^2 \to \mathbb{R}^2$.
   (a) T rotates points (counterclockwise) about the origin by $\frac{\pi}{3}$ radians and then reflects them about the y-axis.
   (b) T rotates points (counterclockwise) about the origin by $\frac{3\pi}{4}$ radians and then projects them onto the y-axis. [𝔸]-354
   (c) T rotates points clockwise about the origin by $\frac{\pi}{6}$ radians and then reflects about the line $y = -x$.
15. Find a sequence of elementary matrices whose product is the matrix derived in exercise 12d. Hint: try a product of the form

$$\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & n \end{bmatrix}\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

16. Reflection about lines other than the axes are linear transformations too and can be realized by a composition of rotations and reflection. To demonstrate, derive the standard matrix for reflection about the line $y = -\dfrac{36}{77}x$ according to the following steps.
   (a) Pick a point (any point) P not on the line.
   (b) Calculate the sine of the angle the line makes with the x-axis, $\sin\left(\tan^{-1}\left(\dfrac{36}{77}\right)\right)$.
   (c) Calculate the cosine of the angle the line makes with the x-axis, $\cos\left(\tan^{-1}\left(\dfrac{36}{77}\right)\right)$.
   (d) Rotate point P counterclockwise about the origin by angle $\tan^{-1}\left(\frac{36}{77}\right)$. HINT: the answers to parts 16b and 16c are useful here.
   (e) Reflect the point from part 16d about the x-axis.
   (f) Rotate the point from part 16e clockwise about the origin by angle $\tan^{-1}\left(\frac{36}{77}\right)$.

   The product of the matrices (that you hopefully used) in parts 16d, 16e, and 16f is the standard matrix. Graph the line and point P and the point from part 16f on the same set of axes to see that you have calculated correctly.

Answers

reflection about x-axis Using F_x as the name for reflection about the x-axis, note that $F_x\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}\right) = F_x\left(\begin{bmatrix} u_1+v_1 \\ u_2+v_2 \end{bmatrix}\right) = \begin{bmatrix} u_1+v_1 \\ -(u_2+v_2) \end{bmatrix} = \begin{bmatrix} u_1+v_1 \\ -u_2-v_2 \end{bmatrix} = \begin{bmatrix} u_1 \\ -u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ -v_2 \end{bmatrix} = F_x\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\right) + F_x\left(\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}\right)$ and $F_x\left(c\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\right) = F_x\left(\begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix}\right) = \begin{bmatrix} cu_1 \\ -cu_2 \end{bmatrix} = c\begin{bmatrix} u_1 \\ -u_2 \end{bmatrix} = cF_x\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\right)$. This shows that $F_x(\mathbf{u}+\mathbf{v}) = F_x(\mathbf{u}) + F_x(\mathbf{v})$ and $F_x(c\mathbf{u}) = cF_x(\mathbf{u})$ for any vectors $\mathbf{u}$ and $\mathbf{v}$ and scalar c, so F_x is linear by definition.

four cases
- rows are multiples of one another: if a and b are nonzero, then $N_{1,:} = \frac{a}{b}N_{2,:}$ and $\frac{b}{a}N_{1,:} = N_{2,:}$
- one of the rows contains two zeros: $a = 0$ or $b = 0$ while k and ℓ are arbitrary
- columns are multiples of one another: if k and ℓ are nonzero, then $N_{:,1} = \frac{k}{\ell}N_{:,2}$ and $\frac{\ell}{k}N_{:,1} = N_{:,2}$
- one of the columns contains two zeros: $k = 0$ or $\ell = 0$ while a and b are arbitrary

4.5 Isomorphisms

The word *vector* has been used in a number of different ways in this book. In section 1.3 the word vector was said to have the understood meaning from physics or calculus and *represented* using the angled bracket notation $\langle x, y\rangle$. $n \times 1$ matrices were called column vectors, or just vectors, and were said to *represent* vectors despite being different objects. The calculus/physics idea of a vector was brought to life geometrically in section 1.4 when a vector was *represented* by an arrow with both magnitude and direction. In section 4.2 it was noted that vectors in a vector space have unique *representations* as linear combinations of basis vectors. Most recently, vectors (with tails at the origin, in section 4.4) were *represented* by points. In all instances, these were *representations* of vectors, not vectors outright. Only in section 4.1, where the word *vector* was used to refer to any element of a vector space, did we have a definition. To be clear, this is the one and only definition of vector. All other uses will have to be justified from within this umbrella.

By definition, $\mathbb{R}^n$ is the set of all ordered lists of n real numbers. It is not, on the surface, a vector space at all. Elements of $\mathbb{R}^n$ are therefore not inherently vectors! It is only once addition and scalar multiplication are defined (and adhere to the ten properties outlined in section 4.1) that $\mathbb{R}^n$ becomes a vector space. When nothing is said to the contrary, addition and scalar multiplication in $\mathbb{R}^n = \{r_1, r_2, \ldots, r_n : r_i \in \mathbb{R}, i = 1, 2, \ldots, n\}$ are understood to be defined element-wise. That is, for any elements $r_1, r_2, \ldots, r_n \in \mathbb{R}^n$ and $s_1, s_2, \ldots, s_n \in \mathbb{R}^n$

$$r_1, r_2, \ldots, r_n \;+\; s_1, s_2, \ldots, s_n \;=\; r_1 + s_1, r_2 + s_2, \ldots, r_n + s_n$$

and for any element $r_1, r_2, \ldots, r_n \in \mathbb{R}^n$ and scalar $c \in \mathbb{R}$,

$$c \;\times\; r_1, r_2, \ldots, r_n \;=\; cr_1, cr_2, \ldots, cr_n.$$

These definitions should remind you of the definitions of matrix addition and scalar multiplication, which are defined entry-wise. For $n \times 1$ matrices, addition and scalar multiplication are defined as follows. For any elements $\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} \in \mathcal{M}_{n\times 1}(\mathbb{R})$ and $\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} \in \mathcal{M}_{n\times 1}(\mathbb{R})$

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} + \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} r_1 + s_1 \\ r_2 + s_2 \\ \vdots \\ r_n + s_n \end{bmatrix}$$

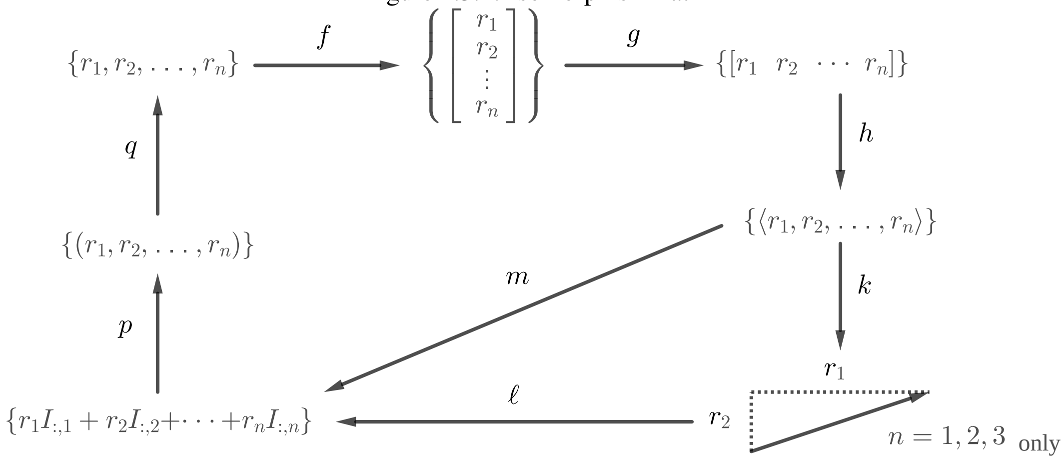
and for any element $\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} \in \mathcal{M}_{n\times 1}(\mathbb{R})$ and scalar $c \in \mathbb{R}$,

$$c \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} = \begin{bmatrix} cr_1 \\ cr_2 \\ \vdots \\ cr_n \end{bmatrix}.$$

So what is the difference between elements of $\mathbb{R}^n$ and elements of $\mathcal{M}_{n\times 1}(\mathbb{R})$? Functionally there is no difference! There is no way to distinguish elements of $\mathbb{R}^n$ and elements of $\mathcal{M}_{n\times 1}(\mathbb{R})$ based purely on their properties. Each ordered list of real numbers $r_1, r_2, \ldots, r_n$ could just as easily be written as a column matrix

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$$

Figure 4.5.1: Isomorphism Path



NOTE: $r_i \in \mathbb{R}, i = 1, 2, \ldots, n$

and vice versa. The sum of two ordered lists of real numbers could just as easily be written as a sum of two column matrices and vice versa. Each scalar multiple of an ordered list of real numbers could just as easily be written as a scalar multiple of a column matrix and vice versa. When two sets are interchangeable in form and function, we say they are isomorphic.

Formally, two sets are **isomorphic** if there exists an isomorphism between them. What defines an isomorphism depends on the structure of the sets. A vector space is a set endowed with two operations. The set defines the elements of the vector space and the operations define the structure. An **isomorphism** between vector spaces maps each element of one vector space to exactly one element of the other without missing any and preserves vector addition and scalar multiplication. Such an isomorphism can be understood as the mathematical formalization allowing the free flow between one representation of a vector and another. It supplies the rigor behind using row vectors, column vectors, ordered lists, arrows, points, linear combinations, and vectors in the sense of calculus or physics as if they were all the same thing. For example, the map $T : \mathbb{R}^n \to \mathcal{M}_{n\times1}(\mathbb{R})$,

$$T(r_1, r_2, \ldots, r_n) = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$$

is an isomorphism. Can you verify this claim? Answer on page 150. To complete the formalism, the following definitions are introduced. A map $T : A \to B$ is called

1. **onto** if for each $b \in B$ the equation $T(a) = b$ has at least one solution $a \in A$.

2. **one-to-one** if for each $b \in B$ the equation $T(a) = b$ has at most one solution $a \in A$.

If A and B are vector spaces, then T is an isomorphism if it is one-to-one, onto, and linear. Being one-to-one and onto assures "each element of one vector space corresponds to exactly one element of the other" and being linear assures it "preserves vector addition and scalar multiplication".

When we use the various representations of elements of $\mathbb{R}^n$ interchangeably, we are relying on the existence of an isomorphism from each one to each other. Much like showing that a list of statements are equivalent by showing a path of implications from any statement to any other, this can be done by showing a path of isomorphisms from any vector space to any other. This is because the composition of isomorphisms is an isomorphism. Can you justify this claim? Answer on page 151. See figure 4.5.1. Once isomorphisms $f, g, h, k, \ell, m, p, q$ between the sets are demonstrated to exist, each vector space is isomorphic to each other by composition. For example, $g : \mathcal{M}_{n\times1}(\mathbb{R}) \to \mathcal{M}_{1\times n}(\mathbb{R})$ defined by

$$g\left(\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}\right) = \begin{bmatrix} r_1 & r_2 & \cdots & r_n \end{bmatrix}$$

is an isomorphism. Can you justify this? Answer on page 151.

Maybe more surprising is the claim that all n-dimensional vector spaces over the real numbers (those defined for real number scalars) are isomorphic. In different words, we might say up to isomorphism, there is only one vector space over the real numbers. Different vector spaces may look different and contain different objects, but they all have the same structure and therefore are interchangeable. This claim can be proven by leaning on the fact that an n-dimensional vector space has a basis with n elements.

Let V and W be n-dimensional vector spaces. By definition, each has a basis with n elements. Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $C = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ be bases for V and W, respectively, and define

$$\begin{aligned} f : V \to \mathbb{R}^n, f(\mathbf{v}) = r_1, r_2, \ldots, r_n \text{ where } \mathbf{v} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_n\mathbf{v}_n \\ g : W \to \mathbb{R}^n, g(\mathbf{w}) = s_1, s_2, \ldots, s_n \text{ where } \mathbf{w} = s_1\mathbf{w}_1 + s_2\mathbf{w}_2 + \cdots + s_n\mathbf{w}_n \end{aligned}$$

Since the expression of an element of a vector space as a linear combination of basis vectors is unique, f and g are well-defined (they are actually functions, not simply relations). Given an arbitrary element $\mathbf{r} = r_1, r_2, \ldots, r_n$ of $\mathbb{R}^n$, let $\mathbf{v} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_n\mathbf{v}_n$. Since vector spaces are closed under linear combinations, $\mathbf{v}$ is in V. Furthermore, $f(\mathbf{v}) = r_1, r_2, \ldots, r_n$ so f is onto. Now suppose there is a second element $\mathbf{u}$ in V such that $f(\mathbf{u}) = r_1, r_2, \ldots, r_n$. By definition of f it must be that $\mathbf{u} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_n\mathbf{v}_n$ so $\mathbf{u} = \mathbf{v}$ and f is one-to-one. Now let $\mathbf{u}$ and $\mathbf{v}$ be arbitrary elements of V, and write $\mathbf{u} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_n\mathbf{v}_n$ and $\mathbf{v} = s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_n\mathbf{v}_n$. Letting c be an arbitrary scalar,

$$\begin{aligned} f(\mathbf{u} + c\mathbf{v}) &= f(r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_n\mathbf{v}_n + c(s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_n\mathbf{v}_n)) \\ &= f(r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_n\mathbf{v}_n + cs_1\mathbf{v}_1 + cs_2\mathbf{v}_2 + \cdots + cs_n\mathbf{v}_n) \\ &= f((r_1 + cs_1)\mathbf{v}_1 + (r_2 + cs_2)\mathbf{v}_2 + \cdots + (r_n + cs_n)\mathbf{v}_n) \\ &= r_1 + cs_1, r_2 + cs_2, \ldots, r_n + cs_n \\ &= r_1, r_2, \ldots, r_n \; + \; cs_1, cs_2, \ldots, cs_n \\ &= r_1, r_2, \ldots, r_n \; + \; c \times s_1, s_2, \ldots, s_n \\ &= f(\mathbf{u}) + cf(\mathbf{v}) \end{aligned}$$

so f is linear. Because f is one-to-one, onto, and linear, f is an isomorphism. By similar argument, g is also an isomorphism. By exercise 13e g^{-1} is an isomorphism. Since the composition of isomorphisms is an isomorphism, $g^{-1} \circ f : V \to W$ is an isomorphism.

Key Concepts

onto a map $T : A \to B$ such that for each $b \in B$ the equation $T(a) = b$ has at least one solution $a \in A$.

one-to-one a map $T : A \to B$ such that for each $b \in B$ the equation $T(a) = b$ has at most one solution $a \in A$.

isomorphism a one-to-one, onto, linear transformation between vector spaces.

isomorphic vector spaces between which there exists an isomorphism.

composition of isomorphisms is an isomorphism.

Exercises

1. Which functions are one-to-one? [$\mathbb{S}$]-312
   (a) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \sin(x)$
   (b) $q : \mathbb{R} \to (0, \infty)$, $q(x) = e^x$
   (c) $p : \mathbb{R} \to [0, \infty)$, $p(x) = x^2$
   (d) $h : [0, \infty) \to \mathbb{R}$, $h(x) = \sqrt{x}$
   (e) $g : \mathbb{R} \to \mathbb{R}$, $g(x) = 3x - 9$
2. Which functions are onto? [$\mathbb{S}$]-312

(a) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = -2x^3$

(b) $q : \mathbb{R} \to \mathbb{R}$, $q(x) = e^x + 1$

(c) $p : [0, \infty) \to \mathbb{R}$, $p(x) = x^2$

(d) $g : \left[0, \frac{\pi}{2}\right] \to \mathbb{R}$, $g(x) = \cos(x)$

(e) $h : [0, \infty) \to [0, \infty)$, $h(x) = \sqrt{x}$

3. Which of the functions in exercise 1 are onto?

4. Which of the functions in exercise 2 are one-to-one?

5. Which of the requirements of isomorphism does $T : \mathbb{R}^3 \to \mathbb{P}_3(\mathbb{R})$, $T\left(\begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix}\right) = (x-r_1)(x-r_2)(x-r_3)$ fail to satisfy?

   (a) T is one-to-one

   (b) T is onto

   (c) T is linear

6. Which of the requirements of isomorphism does $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^4$, $T(\langle s_1, s_2, s_3, s_4, s_5, \ldots \rangle) = s_1, s_2, s_3, s_4$ fail to satisfy? [𝕊]-313

   (a) T is one-to-one

   (b) T is onto

   (c) T is linear

7. Which of the requirements of isomorphism does $T : \mathbb{C} \to \mathcal{M}_{2\times 2}(\mathbb{R})$, $T(a + bi) = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ fail to satisfy?

   (a) T is one-to-one

   (b) T is onto

   (c) T is linear

8. Is the transformation $\det : \mathcal{M}_{2\times 2}(\mathbb{R}) \to \mathbb{R}$

   (a) linear?

   (b) one-to-one?

   (c) onto?

   (d) an isomorphism?

9. Is the transformation $T : C([0,1]) \to C([0,1])$, $T(f) = e^x f(x)$ [𝕊]-313

   (a) linear?

   (b) one-to-one?

   (c) onto?

   (d) an isomorphism?

10. Is the transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ (i) one-to-one? (ii) onto? (iii) an isomorphism?

    (a) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 3x - 4y \\ -x + 3y \\ 2x + 2y \end{bmatrix}$

    (b) $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 2x - 3y + 7z \\ -x + y - 2z \end{bmatrix}$

    (c) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 14 & -1 \\ -15 & 15 \\ 12 & 9 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$ [𝕊]-313

    (d) $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 13 & -15 & -9 \\ 0 & -3 & 10 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}$

    (e) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -5 & 11 \\ -14 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$ [𝕊]-354

    (f) $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 5 & -9 \\ -15 & 27 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$

11. For what type of matrix M is $T_M : \mathbb{R}^n \to \mathbb{R}^m$ an isomorphism?

12. It is claimed that the vector spaces in figure 4.5.1 are all isomorphic, and a formula for isomorphism f is provided in the text. Provide a formula for the isomorphism

    (a) g

    (b) h [𝔸]-354

    (c) m

    (d) p [𝔸]-354

    (e) q

13. Justify the claim.

    (a) the statement "$T_M : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one" may be added to the list of equivalent statements of theorem 5. [𝔸]-354

    (b) the statement "$T_M : \mathbb{R}^n \to \mathbb{R}^m$ is onto" may be added to the list of equivalent statements of theorem 6.

    (c) the statements "$T_M : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one" and "$T_M : \mathbb{R}^n \to \mathbb{R}^m$ is onto" may be added to the list of equivalent statements of theorem 7. [𝔸]-354

    (d) If $f : V \to W$ is an isomorphism between vector spaces V and W, then f is invertible.

    (e) If $f : V \to W$ is an isomorphism from vector space V to vector space W, then f^{-1} is an isomorphism. [𝔸]-354

Answers

first isomorphism An isomorphism between vector spaces maps each element of one vector space to exactly one element of the other without missing any and preserves vector addition and scalar multiplication. The map

$T : \mathbb{R}^n \to \mathcal{M}_{n\times 1}(\mathbb{R})$,

$$T(r_1, r_2, \ldots, r_n) = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$$

does just that because

1. the *arbitrary* element $r_1, r_2, \ldots, r_n \in \mathbb{R}^n$ maps via T to (the specific element) $\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} \in \mathcal{M}_{n\times 1}(\mathbb{R})$ and only $\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$, so T maps each element of $\mathbb{R}^n$ to exactly one element of $\mathcal{M}_{n\times 1}(\mathbb{R})$.
2. (the specific element) $r_1, r_2, \ldots, r_n \in \mathbb{R}^n$ maps via T to the *arbitrary* element $\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} \in \mathcal{M}_{n\times 1}(\mathbb{R})$, so T does not miss any elements of $\mathcal{M}_{n\times 1}(\mathbb{R})$.
3. $T(r_1, r_2, \ldots, r_n + s_1, s_2, \ldots, s_n) = T(r_1 + s_1, r_2 + s_2, \ldots, r_n + s_n) = \begin{bmatrix} r_1 + s_1 \\ r_2 + s_2 \\ \vdots \\ r_n + s_n \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} + \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = T(r_1, r_2, \ldots, r_n) + T(s_1, s_2, \ldots, s_n)$, so addition is preserved under T.
4. $T(c \times r_1, r_2, \ldots, r_n) = T(cr_1, cr_2, \ldots, cr_n) = \begin{bmatrix} cr_1 \\ cr_2 \\ \vdots \\ cr_n \end{bmatrix} = c \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} = cT(r_1, r_2, \ldots, r_n)$, so scalar multiplication is preserved under T.

composition of isomorphisms Let A, B, C be vector spaces and $T : A \to B$ and $S : B \to C$ be isomorphisms. We need to show that $S \circ T : A \to C$ is an isomorphism.

1. Let c be an element of C. Then because S is onto, there is at least one $b \in B$ such that $S(b) = c$. Let b be such a solution. Because T is onto, there is at least one $a \in A$ such that $T(a) = b$. Let a be such a solution. Then $S \circ T(a) = S(T(a)) = S(b) = c$ so $S \circ T(a) = c$ has at least one soution and $S \circ T$ is onto. Generally, this shows that **the composition of onto mappings is an onto mapping**.
2. Suppose $S \circ T(a_1) = c$ and $S \circ T(a_2) = c$. Equivalently $S(T(a_1)) = c$ and $S(T(a_2)) = c$. But S is one-to-one, so the equation $S(b) = c$ has at most one solution. Therefore, $T(a_1) = T(a_2) = b$ for the same $b \in B$. Since T is one-to-one, the equation $T(a) = b$ has at most one solution. Therefore $a_1 = a_2$, which shows that for each $c \in C$, the equation $S \circ T(a_1) = c$ has at most one solution and $S \circ T$ is one-to-one. Generally, this shows that **the composition of one-to-one mappings is a one-to-one mapping**.
3. In exercise 24c of section 4.3 you are asked to show that the composition of linear transformations is linear. This completes the proof.

isomorphism g Let $g : \mathcal{M}_{n\times 1}(\mathbb{R}) \to \mathcal{M}_{1\times n}(\mathbb{R})$ be defined by

$$g\left(\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}\right) = \begin{bmatrix} r_1 & r_2 & \cdots & r_n \end{bmatrix}.$$

Then

1. Given $\begin{bmatrix} r_1 & r_2 & \cdots & r_n \end{bmatrix}$ in $\mathcal{M}_{1\times n}(\mathbb{R})$,

$$g\left(\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}\right) = \begin{bmatrix} r_1 & r_2 & \cdots & r_n \end{bmatrix}$$

so g is onto.

2. Given $r = \begin{bmatrix} r_1 & r_2 & \cdots & r_n \end{bmatrix}$ in $\mathcal{M}_{1\times n}(\mathbb{R})$, suppose $g(u) = r$ and $g(v) = r$. Then

$$u = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} = v$$

so g is one-to-one.

3. Let $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$ and $y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T$ be in $\mathcal{M}_{n\times 1}(\mathbb{R})$ and c be a scalar. Using the result of exercise 24d of section 4.3, the following calculation shows the linearity of L.

$$\begin{aligned} L(x + cy) &= L\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + c\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}\right) \\ &= L\left(\begin{bmatrix} x_1 + cy_1 \\ x_2 + cy_2 \\ \vdots \\ x_n + cy_n \end{bmatrix}\right) \\ &= \begin{bmatrix} x_1 + cy_1 & x_2 + cy_2 & \cdots & x_n + cy_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} + c\begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \\ &= L(x) + cL(y) \end{aligned}$$

4.6 Inner Product Spaces

Section 4.5 ended by showing that all n-dimensional vector spaces are isomorphic. That means, for example, $\mathbb{R}^6$, $\mathbb{P}_5(\mathbb{R})$, $\mathcal{M}_{2\times 3}(\mathbb{R})$, and the vector space generated by $G = \{\sin t, \cos t, \sin 2t, \cos 2t, \sin 3t, \cos 3t\}$ are all isomorphic. The essential ingredient of elements of each space is an ordered list of six real numbers. Elements of any of these vector spaces can be represented by such a list. Addition and scalar multiplication in one can be accomplished just as well in another. Representation, addition, and scalar multiplication are not enough to distinguish one from the other. Yet it might be nice to be able to draw distinctions between them. After all, they are different types of objects. Polynomials and elements of G are functions. A list of real numbers is just a list. A matrix is yet a different type of object.

One important feature of $\mathbb{R}^n$ not included in the definition of a vector space is the dot product. We have seen that the dot product allows us to define the magnitudes and orthogonality of elements of $\mathbb{R}^n$. If we could extend the idea of the dot product to any vector space in such a way that it may not be preserved under one-to-one and onto maps, it might prove a useful distinguishing feature. To do that, though, we would need to write down the essential features of the dot product and hope that, as an abstract list of requirements, other vector spaces would admit similar operators. We explore some properties of the dot product presently.

First, representing elements of $\mathbb{R}^n$ by column vectors, the dot product of $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is defined by $\mathbf{u}^T\mathbf{v}$ (see section 1.3). Letting $\mathbf{u} = u_1, u_2, \ldots, u_n$ and $\mathbf{v} = v_1, v_2, \ldots, v_n$, the dot product of $\mathbf{u}$ and $\mathbf{v}$, which we will begin to denote $\mathbf{u} \cdot \mathbf{v}$, is given by

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n. \tag{4.6.1}$$

This is not a new definition, just a new notation and a general formula for computing the dot product. Also in section 1.3 the magnitude of a vector $\mathbf{u}$ is defined by $\sqrt{\mathbf{u}^T\mathbf{u}}$. Using the new notation and formula (4.6.1),

$$\begin{aligned} \|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} &= \sqrt{u_1u_1 + u_2u_2 + \cdots + u_nu_n} \\ &= \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}, \end{aligned}$$

a formula that only makes sense as a magnitude since $\mathbf{u}\cdot\mathbf{u}$ is nonnegative (the squares of real numbers are nonnegative and the sum of nonnegative numbers is nonnegative). If $\mathbf{u} \cdot \mathbf{u}$ were sometimes negative it would not make a good quantity for defining vector magnitude. As such, the nonnegativity of $\mathbf{u}\cdot\mathbf{u}$ is an important property of the dot product.

Second, in section 1.4 it was shown that $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$, meaning that the dot product is commutative. Commutativity (or lack thereof) is a fundamental property of any operator. Since the dot product is commutative any abstraction of the dot product should be commutative too. Note that both nonnegativity (of $\mathbf{u}\cdot\mathbf{u}$) and commutativity (of $\mathbf{u}\cdot\mathbf{v}$) follow directly from properties of the real numbers.

Two other properties, distributivity and a type of associativity, of the dot product follow directly from properties of the real numbers as well. You may have explored these properties in exercises 4 and 8 of section 1.3. To recap using the new notation,

$$(\mathbf{u} + \mathbf{w}) \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{v}$$

and

$$(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}).$$

Can you show these identities are true using properties of real numbers? Answer on page 157. You might think of these properties as a kind of linearity. Since they follow as a natural consequence of using real number scalars, any abstraction of the dot product should have a similar distributive and associative properties.

Fifth, a property that cannot be proven from the previous four but is nonetheless a direct consequence of properties of the real numbers is that

$$\mathbf{u} \cdot \mathbf{u} = 0 \text{ if and only if } \mathbf{u} = \mathbf{0}.$$

That is, the statements "$\mathbf{u} \cdot \mathbf{u} = 0$" and "$\mathbf{u} = \mathbf{0}$" are equivalent. The dot product can be used to determine whether a vector is the zero vector. Can you argue using properties of real numbers that the equivalence is true? Answer on page 157.

These five features of the dot product form the foundation for a useful abstraction. We define an **inner product** on a real vector space V to be any operator $\langle , \rangle : V \times V \to \mathbb{R}$ such that

1. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ for all $\mathbf{u}$ in V (nonnegativity)

2. $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$ (zero vector identity)
3. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ for all $\mathbf{u}, \mathbf{v}$ in V (commutativity)
4. $\langle \mathbf{u} + \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V (preservation of addition)
5. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$ for all $\mathbf{u}, \mathbf{v}$ in V and all scalars c (preservation of scalar multiplication)

The dot product on $\mathbb{R}^n$ as defined in section 1.3 is the canonical, and motivating, example of an inner product. Any real vector space on which an inner product is defined is called an **inner product space**.

Extending the ideas of magnitude, distance, and orthogonality to an arbitrary inner product space is a simple matter as the following chart suggests.

	in $\mathbb{R}^n$	in an n-dimensional inner product space
norm[a]	$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$	$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$
distance[b]	$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$	$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$
orthogonality[c]	$\mathbf{u} \cdot \mathbf{v} = 0$	$\langle \mathbf{u}, \mathbf{v} \rangle = 0$

[a]replacement for the word *magnitude* in an arbitrary inner product space.
[b]see exercise 2 in section 1.4
[c]calculation (1.4.1) proceeds identically if each dot product is replced by an inner product.

We use the words norm instead of magnitude and orthogonal instead of perpendicular because magnitude and perpendicular have special, visual, geometric meaning in $\mathbb{R}^2$ and $\mathbb{R}^3$ that does not readily transer to other contexts. Objects such as matrices and functions, and even vectors in $\mathbb{R}^n$ for $n > 3$ cannot be visualized the same way. What would it mean for two polynomials to be perpendicular, for example? What would be the geometric magnitude of a matrix? The familiar geometric notions of magnitude and perpendicularity of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ have no geometric analogy in other vector spaces such as $\mathbb{P}_2(\mathbb{R})$ and $\mathcal{M}_{m\times n}(\mathbb{R})$. No matter. This is really the purpose of the above chart. Orthogonality (the abstraction of perpendicularity) of two vectors is *defined by* the requirement that their inner product be zero whatever that may look like geometrically. Therein lies the power of mathematical abstraction. A notion such as perpendicularity, which we can clearly see and grasp in $\mathbb{R}^2$, can be extended to other sets where no analogous picture can be drawn. Likewise the norm (abstraction of magnitude) of a matrix M is *defined by* the quantity $\sqrt{\langle M, M \rangle}$ whether it has geometric meaning or not.

Can you verify that $\langle , \rangle : \mathbb{P}_2(\mathbb{R}) \times \mathbb{P}_2(\mathbb{R}) \to \mathbb{R}$,

$$\langle p_0 + p_1 x + p_2 x^2, q_0 + q_1 x + q_2 x^2 \rangle = \frac{2}{5}p_2 q_2 + \frac{2}{3}p_0 q_2 + \frac{2}{3}p_1 q_1 + \frac{2}{3}p_2 q_0 + 2p_0 q_0 \tag{4.6.2}$$

is an inner product? If you have taken calculus, this is equivalent to $\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)\,dx$. Answer without using calculus on page 157.

Key Concepts

inner product an operator $\langle , \rangle : V \times V \to \mathbb{R}$ on a real vector space V such that

1. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ for all $\mathbf{u}$ in V
2. $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$
3. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ for all $\mathbf{u}, \mathbf{v}$ in V
4. $\langle \mathbf{u} + \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V
5. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$ for all $\mathbf{u}, \mathbf{v}$ in V and all scalars c

inner product space a vector space endowed with an inner product.

norm extension of the idea of magnitude in $\mathbb{R}^2$ or $\mathbb{R}^3$ to vectors in any inner product space. The norm of a vector $\mathbf{v}$ is denoted $\|\mathbf{v}\|$ and is calculated as $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

distance extension of the idea of distance in $\mathbb{R}^2$ or $\mathbb{R}^3$ to vectors in any inner product space. The distance between two vectors $\mathbf{u}$ and $\mathbf{v}$ is denoted $d(\mathbf{u}, \mathbf{v})$ and is calculated as $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.

orthogonal extension of the idea of perpendicular in $\mathbb{R}^2$ or $\mathbb{R}^3$ to vectors in any inner product space. Two vectors $\mathbf{u}$ and $\mathbf{v}$ are said to be orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Exercises

1. Verify that the operator is an inner product.

   (a) $\langle , \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$,
   $$\left\langle \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\rangle = 2u_1v_1 + 3u_2v_2$$
   [$\mathbb{S}$]-314

   (b) $\langle , \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$,
   $$\left\langle \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\rangle = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

   (c) $\langle , \rangle : \mathbb{P}_2(\mathbb{R}) \times \mathbb{P}_2(\mathbb{R}) \to \mathbb{R}$,
   $$\langle p, q \rangle = p(0)q(0) + p(1)q(1) + p(2)q(2)$$

   (d) $\langle , \rangle : \mathbb{P}_2(\mathbb{R}) \times \mathbb{P}_2(\mathbb{R}) \to \mathbb{R}$,
   $$\langle p, q \rangle = p(0)q(0) + p(1)q(1) + p(2)q(2) + p(3)q(3)$$

   (e) $\langle , \rangle : C([0,1]) \times C([0,1]) \to \mathbb{R}$,
   $$\langle f, g \rangle = \int_0^1 f(x)g(x)\,dx$$

   (f) $\langle , \rangle : C([0,2\pi]) \times C([0,2\pi]) \to \mathbb{R}$,
   $$\langle f, g \rangle = \frac{1}{\pi}\int_0^{2\pi} f(x)g(x)\,dx$$

   (g) $\langle , \rangle : \mathcal{M}_{2\times 2}(\mathbb{R}) \times \mathcal{M}_{2\times 2}(\mathbb{R}) \to \mathbb{R}$,
   $$\langle M, N \rangle = (MN^T)_{1,1} + (MN^T)_{2,2}$$

2. Using the inner product of question 1b, calculate

   (a) $\left\langle \begin{bmatrix} -6 & -1 \end{bmatrix}^T, \begin{bmatrix} -2 & 6 \end{bmatrix}^T \right\rangle$

   (b) $\left\langle \begin{bmatrix} -1 & 1 \end{bmatrix}^T, \begin{bmatrix} 2 & 0 \end{bmatrix}^T \right\rangle$

   (c) $\left\langle \begin{bmatrix} 6 & -3 \end{bmatrix}^T, \begin{bmatrix} 3 & -5 \end{bmatrix}^T \right\rangle$ [$\mathbb{A}$]-355

   (d) $\left\| \begin{bmatrix} 1 & -1 \end{bmatrix}^T \right\|$

   (e) $\left\| \begin{bmatrix} 2 & 3 \end{bmatrix}^T \right\|$

   (f) $\left\| \begin{bmatrix} -1 & 6 \end{bmatrix}^T \right\|$ [$\mathbb{A}$]-355

3. Using the inner product of question 1d, calculate

   (a) $\left\langle 2 + 5x + 6x^2, -2 - 3x - x^2 \right\rangle$ [$\mathbb{A}$]-355

   (b) $\left\langle -3 + x^2, 1 - 2x + 4x^2 \right\rangle$

   (c) $\left\langle 3 - 4x - 3x^2, -2 + 1 - 5x^2 \right\rangle$

   (d) $\left\| 2 - 2x + x^2 \right\|$ [$\mathbb{A}$]-355

   (e) $\|-4 - x\|$

   (f) $\left\| -5 + 4x - 3x^2 \right\|$

4. Using the inner product of question 1f, calculate

   (a) $\left\langle x^2, 2 - x \right\rangle$

   (b) $\left\langle x + 10, \frac{1}{x-10} \right\rangle$ [$\mathbb{A}$]-355

   (c) $\langle \cos x, \sin x \rangle$

   (d) $\|\cos x\|$ [$\mathbb{A}$]-355

   (e) $\|x\|$

   (f) $\left\| 5(1 + x - x^2) \right\|$

5. Using the inner product of question 1g, calculate

   (a) $\left\langle \begin{bmatrix} -9 & 0 \\ -8 & 7 \end{bmatrix}, \begin{bmatrix} 5 & 6 \\ -2 & -3 \end{bmatrix} \right\rangle$ [$\mathbb{A}$]-355

   (b) $\left\langle \begin{bmatrix} -8 & 4 \\ 9 & -9 \end{bmatrix}, \begin{bmatrix} -9 & 4 \\ 7 & 0 \end{bmatrix} \right\rangle$

   (c) $\left\langle \begin{bmatrix} 3 & 9 \\ 4 & -3 \end{bmatrix}, \begin{bmatrix} 0 & 6 \\ 8 & -9 \end{bmatrix} \right\rangle$

   (d) $\left\| \begin{bmatrix} -1 & 7 \\ -3 & 9 \end{bmatrix} \right\|$ [$\mathbb{A}$]-355

   (e) $\left\| \begin{bmatrix} 9 & -9 \\ -4 & -2 \end{bmatrix} \right\|$

   (f) $\left\| \begin{bmatrix} 0 & -3 \\ -9 & 7 \end{bmatrix} \right\|$

6. Using the inner product of question 1a, find the distance between

   (a) $\begin{bmatrix} -12 & 2 \end{bmatrix}^T$ and $\begin{bmatrix} -13 & 5 \end{bmatrix}^T$

   (b) $\begin{bmatrix} -8 & 11 \end{bmatrix}^T$ and $\begin{bmatrix} -12 & 13 \end{bmatrix}^T$

   (c) $\begin{bmatrix} 9 & -6 \end{bmatrix}^T$ and $\begin{bmatrix} -5 & -9 \end{bmatrix}^T$

   (d) $\begin{bmatrix} 8 & -4 \end{bmatrix}^T$ and $\begin{bmatrix} 14 & 3 \end{bmatrix}^T$ [$\mathbb{A}$]-355

7. Using the inner product of question 1c, find the distance between

   (a) $3 - 3x + 4x^2$ and $-5 + 6x + 5x^2$

   (b) $4x - 3x^2$ and $-4 - 5x - 6x^2$

   (c) $1 + 3x - 2x^2$ and $6 + 5x - x^2$

   (d) $6 + 2x - 6x^2$ and $-5 - 3x - 2x^2$ [$\mathbb{A}$]-355

8. Using the inner product of question 1e, find the distance between

   (a) x and x^2 [$\mathbb{A}$]-355

   (b) x and e^x

   (c) $\sin \pi x$ and $\cos \pi x$

   (d) $2x$ and $\frac{1}{x^2+1}$

9. Using the inner product of question 1g, find the distance between

   (a) $\begin{bmatrix} 3 & -5 \\ -3 & 1 \end{bmatrix}$ and $\begin{bmatrix} -8 & -5 \\ 9 & -9 \end{bmatrix}$

   (b) $\begin{bmatrix} -4 & -5 \\ -7 & -6 \end{bmatrix}$ and $\begin{bmatrix} -5 & 8 \\ -8 & 9 \end{bmatrix}$

   (c) $\begin{bmatrix} -8 & 9 \\ 4 & -5 \end{bmatrix}$ and $\begin{bmatrix} 2 & 3 \\ -4 & -7 \end{bmatrix}$

   (d) $\begin{bmatrix} 9 & -3 \\ -6 & -9 \end{bmatrix}$ and $\begin{bmatrix} -3 & 3 \\ 2 & 9 \end{bmatrix}$ [$\mathbb{A}$]-355

10. In $\mathbb{R}^2$ with the inner product of question 1b, are $\mathbf{u}$ and $\mathbf{v}$ orthogonal?
    (a) $\mathbf{u} = \begin{bmatrix} 3 & 10 \end{bmatrix}^T$ and $\mathbf{v} = \begin{bmatrix} 10 & -5 \end{bmatrix}^T$ [𝕊]-314
    (b) $\mathbf{u} = \begin{bmatrix} -1 & -10 \end{bmatrix}^T$ and $\mathbf{v} = \begin{bmatrix} -6 & 1 \end{bmatrix}^T$
    (c) $\mathbf{u} = \begin{bmatrix} 9 & 5 \end{bmatrix}^T$ and $\mathbf{v} = \begin{bmatrix} 7 & -20 \end{bmatrix}^T$
    (d) $\mathbf{u} = \begin{bmatrix} -8 & 4 \end{bmatrix}^T$ and $\mathbf{v} = \begin{bmatrix} -3 & -10 \end{bmatrix}^T$
    (e) $\mathbf{u} = \begin{bmatrix} 6 & 4 \end{bmatrix}^T$ and $\mathbf{v} = \begin{bmatrix} -3 & 5 \end{bmatrix}^T$
11. In $\mathbb{P}_2(\mathbb{R})$ with the inner product of question 1d, are $\mathbf{u}$ and $\mathbf{v}$ orthogonal?
    (a) $\mathbf{u} = x(x-1)$ and $\mathbf{v} = (x-2)(x-3)$
    (b) $\mathbf{u} = x(x-2)$ and $\mathbf{v} = (x+1)(x-3)$
    (c) $\mathbf{u} = (x-1)(x-3)$ and $\mathbf{v} = x^2 - 2x$ [𝕊]-314
    (d) $\mathbf{u} = 2x^2 - 6x$ and $\mathbf{v} = x^2 - 3x + 2$
    (e) $\mathbf{u} = x^2 - 2x - 3$ and $\mathbf{v} = 3x^2 - 3$
12. In $C([0, 2\pi])$ with the inner product of question 1f, are $\mathbf{u}$ and $\mathbf{v}$ orthogonal?
    (a) $\mathbf{u} = \sin x$ and $\mathbf{v} = \cos x$ [𝕊]-314
    (b) $\mathbf{u} = 3$ and $\mathbf{v} = 2 - \frac{2}{\pi}x$
    (c) $\mathbf{u} = \sin x$ and $\mathbf{v} = \sin 2x$
    (d) $\mathbf{u} = e^x$ and $\mathbf{v} = -x$
    (e) $\mathbf{u} = \frac{4}{4x+1}$ and $\mathbf{v} = x - \frac{23}{12}$
13. In $\mathcal{M}_{2\times 2}(\mathbb{R})$ with the inner product of question 1g, are $\mathbf{u}$ and $\mathbf{v}$ orthogonal?
    (a) $\mathbf{u} = \begin{bmatrix} 2 & 7 \\ -2 & 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -2 & 8 \\ -6 & -4 \end{bmatrix}$ [𝕊]-314
    (b) $\mathbf{u} = \begin{bmatrix} -7 & 7 \\ 1 & 4 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 8 & -5 \\ -9 & 2 \end{bmatrix}$
    (c) $\mathbf{u} = \begin{bmatrix} -8 & -5 \\ -1 & 8 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -3 & -6 \\ 6 & -6 \end{bmatrix}$
    (d) $\mathbf{u} = \begin{bmatrix} -6 & 5 \\ 2 & -2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 & 6 \\ -7 & -4 \end{bmatrix}$
    (e) $\mathbf{u} = \begin{bmatrix} -5 & -4 \\ 5 & -2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 & -1 \\ 5 & 7 \end{bmatrix}$
14. Describe all vectors orthogonal to
    (a) $\begin{bmatrix} 3 & 5 \end{bmatrix}^T$ in $\mathbb{R}^2$ with the inner product of question 1a. [𝔸]-355
    (b) $x(x-5)$ in $\mathbb{P}_2(\mathbb{R})$ with the inner product of question 1c.
    (c) $f(x) = 1$ in $C([0, 1])$ with the inner product of question 1e. [𝔸]-355
    (d) $\begin{bmatrix} 8 & 0 \\ 1 & -3 \end{bmatrix}$ in $\mathcal{M}_{2\times 2}(\mathbb{R})$ with the inner product of question 1g.
15. Describe all vectors orthogonal to
    (a) $\begin{bmatrix} 3 & 5 \end{bmatrix}^T$ in $\mathbb{R}^2$ with the inner product of question 1b.
    (b) $x(x-2)(x+3)$ in $\mathbb{P}_2(\mathbb{R})$ with the inner product of question 1d.
16. Suppose for vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ of an inner product space, $\langle \mathbf{u}, \mathbf{v} \rangle = 3$ and $\langle \mathbf{u}, \mathbf{w} \rangle = \frac{1}{3}$. Use this information to compute
    (a) $\langle \mathbf{u}, 3\mathbf{v} \rangle$
    (b) $\langle \mathbf{u}, \mathbf{v} + 2\mathbf{w} \rangle$ [𝔸]-355
    (c) $\langle -2\mathbf{v}, \mathbf{u} \rangle$
    (d) $\langle 3\mathbf{w}, 2\mathbf{u} \rangle$ [𝔸]-355
17. For what values of a and b is the operator $\langle , \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$,
$$\left\langle \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\rangle = a u_1 v_1 + b u_2 v_2$$
an inner product?
18. Explain why $\langle , \rangle : \mathbb{P}_2(\mathbb{R}) \times \mathbb{P}_2(\mathbb{R}) \to \mathbb{R}$,
$$\langle p, q \rangle = p(0)q(0) + p(1)q(1)$$
is not an inner product. [𝔸]-355
19. Justify the claim.
    (a) For any vector $\mathbf{v}$ in an inner product space, $\langle \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{0} \rangle = 0$. [𝕊]-315
    (b) For any scalar c and any vectors $\mathbf{u}$ and $\mathbf{v}$ of an inner product space, $\langle \mathbf{u}, c\mathbf{v} \rangle = c \langle \mathbf{u}, \mathbf{v} \rangle$.
    (c) For any any vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ of an inner product space, $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$. [𝕊]-315
    (d) For any vectors $\mathbf{u}$ and $\mathbf{v}$ of an inner product space, $\|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2 = 4 \langle \mathbf{u}, \mathbf{v} \rangle$.
    (e) For any orthogonal vectors $\mathbf{u}$ and $\mathbf{v}$ of an inner product space, $\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Answers

dot product identities Let $\mathbf{u} = u_1, u_2, \ldots, u_n$, $\mathbf{v} = v_1, v_2, \ldots, v_n$, $\mathbf{w} = w_1, w_2, \ldots, w_n$ be arbitrary elements of $\mathbb{R}^n$. Then

$$\begin{aligned}
(\mathbf{u} + \mathbf{w}) \cdot \mathbf{v} &= (u_1, u_2, \ldots, u_n \ + \ w_1, w_2, \ldots, w_n) \ \cdot \ v_1, v_2, \ldots, v_n \\
&= u_1 + w_1, u_2 + w_2, \ldots, u_n + w_n \ \cdot \ v_1, v_2, \ldots, v_n \\
&= (u_1 + w_1)v_1 + (u_2 + w_2)v_2 + \cdots + (u_n + w_n)v_n \\
&= u_1v_1 + w_1v_1 + u_2v_2 + w_2v_2 + \cdots + u_nv_n + w_nv_n \\
&= (u_1v_1 + u_2v_2 + \cdots + u_nv_n) + (w_1v_1 + w_2v_2 + \cdots + w_nv_n) \\
&= \mathbf{u} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{v}
\end{aligned}$$

and

$$\begin{aligned}
(c\mathbf{u}) \cdot \mathbf{v} &= (c \ \times \ u_1, u_2, \ldots, u_n) \ \cdot \ v_1, v_2, \ldots, v_n \\
&= (cu_1, cu_2, \ldots, cu_n) \ \cdot \ v_1, v_2, \ldots, v_n \\
&= cu_1v_1 + cu_2v_2 + \cdots + cu_nv_n \\
&= c(u_1v_1 + u_2v_2 + \cdots + u_nv_n) \\
&= c(\mathbf{u} \cdot \mathbf{v}).
\end{aligned}$$

dot product zero Suppose $\mathbf{u} \cdot \mathbf{u} = \mathbf{0}$ for some vector $\mathbf{u} = u_1, u_2, \ldots, u_n$ in $\mathbb{R}^n$. That is, $u_1^2 + u_2^2 + \cdots + u_n^2 = 0$. Since this is a sum of nonnegative real numbers that add to zero, each term must itself be zero: $u_1 = u_2 = \cdots = u_n = 0$. Hence $\mathbf{u} = \mathbf{0}$. Now suppose $\mathbf{u} = \mathbf{0}$. Then $u_1 = u_2 = \cdots = u_n = 0$ and $\mathbf{u} \cdot \mathbf{u} = u_1^2 + u_2^2 + \cdots + u_n^2 = 0 + 0 + \cdots + 0 = 0$.

inner product on polynomials Given polynomials $p(x) = p_0 + p_1x + p_2x^2$, $q(x) = q_0 + q_1x + q_2x^2$, and $r(x) = r_0 + r_1x + r_2x^2$ in $\mathbb{P}_2(\mathbb{R})$ and the operator

$$\langle p(x), q(x) \rangle = \frac{2}{5}p_2q_2 + \frac{2}{3}p_0q_2 + \frac{2}{3}p_1q_1 + \frac{2}{3}p_2q_0 + 2p_0q_0,$$

the most challenging part of the verification is the need for some fancy algebra to show properties 1 and 2. The expression $\langle p, p \rangle$ is manipulated into a sum of squares for this purpose.

1. For any p in $\mathbb{P}_2(\mathbb{R})$,

$$\begin{aligned}
\langle p, p \rangle &= \frac{2}{5}p_2^2 + \frac{4}{3}p_0p_2 + \frac{2}{3}p_1^2 + 2p_0^2 \\
&= \frac{1}{15}\left[6p_2^2 + 20p_0p_2 + 10p_1^2 + 30p_0^2\right] \\
&= \frac{1}{15}\left[(2p_2 + 5p_0)^2 + 2p_2^2 + 5p_0^2 + 10p_1^2\right]
\end{aligned}$$

   which is a sum of squares and therefore greater than or equal to 0.

2. $\langle p, p \rangle = \frac{1}{15}\left[(2p_2 + 5p_0)^2 + 2p_2^2 + 5p_0^2 + 10p_1^2\right]$ equals 0 if and only if $p_0 = p_1 = p_2 = 0$ since the square of each appears as a term in the sum. And $p_0 = p_1 = p_2 = 0$ if and only if $p(x) = 0$ (that is, $p = \mathbf{0}$).

3. For any p, q in $\mathbb{P}_2(\mathbb{R})$,

$$\begin{aligned}
\langle p, q \rangle &= \frac{2}{5}p_2q_2 + \frac{2}{3}p_0q_2 + \frac{2}{3}p_1q_1 + \frac{2}{3}p_2q_0 + 2p_0q_0 \\
&= \frac{2}{5}q_2p_2 + \frac{2}{3}q_2p_0 + \frac{2}{3}q_1p_1 + \frac{2}{3}q_0p_2 + 2q_0p_0 \\
&= \frac{2}{5}q_2p_2 + \frac{2}{3}q_0p_2 + \frac{2}{3}q_1p_1 + \frac{2}{3}q_2p_0 + 2q_0p_0 \\
&= \langle q, p \rangle
\end{aligned}$$

4. For any p, q, r in $\mathbb{P}_2(\mathbb{R})$,

$$\begin{aligned}
\langle p+q, r\rangle &= \left\langle \left(p_0 + p_1 x + p_2 x^2\right) + \left(q_0 + q_1 x + q_2 x^2\right)\ ,\ r_0 + r_1 x + r_2 x^2 \right\rangle \\
&= \left\langle (p_0 + q_0) + (p_1 + q_1)\, x + (p_2 + q_2)\, x^2\ ,\ r_0 + r_1 x + r_2 x^2 \right\rangle \\
&= \frac{2}{5}(p_2 + q_2)\, r_2 + \frac{2}{3}(p_0 + q_0)\, r_2 + \frac{2}{3}(p_1 + q_1)\, r_1 + \frac{2}{3}(p_2 + q_2)\, r_0 + 2\,(p_0 + q_0)\, r_0 \\
&= \frac{2}{5} p_2 r_2 + \frac{2}{5} q_2 r_2 + \frac{2}{3} p_0 r_2 + \frac{2}{3} q_0 r_2 + \frac{2}{3} p_1 r_1 + \frac{2}{3} q_1 r_1 \\
&\quad + \frac{2}{3} p_2 r_0 + \frac{2}{3} q_2 r_0 + 2 p_0 r_0 + 2 q_0 r_0 \\
&= \frac{2}{5} p_2 r_2 + \frac{2}{3} p_0 r_2 + \frac{2}{3} p_1 r_1 + \frac{2}{3} p_2 r_0 + 2 p_0 r_0 \\
&\quad + \frac{2}{5} q_2 r_2 + \frac{2}{3} q_0 r_2 + \frac{2}{3} q_1 r_1 + \frac{2}{3} q_2 r_0 + 2 q_0 r_0 \\
&= \langle p, r\rangle + \langle q, r\rangle
\end{aligned}$$

5. For any p, q in $\mathbb{P}_2(\mathbb{R})$ and scalar c,

$$\begin{aligned}
\langle cp, q\rangle &= \left\langle c\left(p_0 + p_1 x + p_2 x^2\right)\ ,\ q_0 + q_1 x + q_2 x^2 \right\rangle \\
&= \left\langle c p_0 + c p_1 x + c p_2 x^2\ ,\ q_0 + q_1 x + q_2 x^2 \right\rangle \\
&= \frac{2}{5} c p_2 q_2 + \frac{2}{3} c p_0 q_2 + \frac{2}{3} c p_1 q_1 + \frac{2}{3} c p_2 q_0 + 2 c p_0 q_0 \\
&= c\left(\frac{2}{5} p_2 q_2 + \frac{2}{3} p_0 q_2 + \frac{2}{3} p_1 q_1 + \frac{2}{3} p_2 q_0 + 2 p_0 q_0\right) \\
&= c\langle p, q\rangle
\end{aligned}$$

Thus $\langle p, q\rangle = \frac{2}{5} p_2 q_2 + \frac{2}{3} p_0 q_2 + \frac{2}{3} p_1 q_1 + \frac{2}{3} p_2 q_0 + 2 p_0 q_0$ satisfies the five properties of an inner product (and is therefore an inner product on $\mathbb{P}_2(\mathbb{R})$.

Chapter 5

Exploring Vector Spaces and Inner Product Spaces

5.1 Solution Spaces [3.3, 3.6, 4.1, 4.2]

Given a coefficient matrix M and a particular vector $\mathbf{b}$, we can use row reduction to determine whether a solution of $M\mathbf{v} = \mathbf{b}$ exists and find it if it does (section 2.2). We even have an efficient way of finding all the solutions when there are more than one (section 3.7 page 111). Being able to do this on a case-by-case basis is good, but a more critical look at the patterns of free and basic variables leads to more complete understanding of solution sets of linear systems.

Observations

Let M be an $m \times n$ matrix and set C as the set of all linear combinations of the columns of M and N as the solution set of $M\mathbf{v} = \mathbf{0}$. That is,

$$\begin{aligned} C &= \{M\mathbf{v} : \mathbf{v} \in \mathbb{R}^n\} \\ N &= \{\mathbf{v} \in \mathbb{R}^n : M\mathbf{v} = \mathbf{0}\}. \end{aligned}$$

Now observe that C and N are vector spaces. For one, C is the collection of all linear combinations of the columns of M. In other words, $C = \operatorname{span}\{M_{:,1}, M_{:,2}, \ldots, M_{:,n}\}$ and is therefore a vector space (see "span is a subspace" on page 123). For the other, we need to check three things (section 4.1 pages 119 and 120):

1. $M\mathbf{0} = \mathbf{0}$ so $\mathbf{0} \in N$.

2. For any $\mathbf{u}$ and $\mathbf{v}$ in N, $M(\mathbf{u} + \mathbf{v}) = M\mathbf{u} + M\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$ so $\mathbf{u} + \mathbf{v}$ is in N.

3. For any $\mathbf{u}$ in N and scalar c, $M(c\mathbf{v}) = c(M\mathbf{v}) = c\mathbf{0} = \mathbf{0}$ so $c\mathbf{v}$ is in N.

The zero vector is in N and N is closed under vector addition and scalar multiplication. Hence N is a vector space.

C is called the **column space** of M and N is called the **null space** of M. The dimension of the column space of M is called the **rank** of M and the dimension of the null space of M is called the **nullity** of M. For any eigenvalue λ of M, the null space of $M - \lambda I$ is called the **eigenspace** of M corresponding to λ.

Implications

Row operations were defined to maintain the solution sets of linear systems. Solutions of a row reduced linear system are solutions of the original linear system. Using the matrix form for a linear system, this means given a particular vector $\mathbf{b}$, $M\mathbf{v} = \mathbf{b}$ has the exact same solution set as $(EM)\mathbf{v} = \mathbf{b}$ for any elementary matrix E. In particular this means the null space of M (the solution set of $M\mathbf{v} = \mathbf{0}$) and the null space of EM (the solution set of $(EM)\mathbf{v} = \mathbf{0}$) are equal for any elementary matrix E. **Row operations do not affect the null space of a matrix**. Stated another way, $\mathbf{v}$ is in the null space of M if and only if $\mathbf{v}$ is in the null space of EM.

To rigorously prove this claim, let M be an $m \times n$ matrix and E be an $n \times n$ elementary matrix. If $\mathbf{v}$ is in the null space of M then $M\mathbf{v} = \mathbf{0}$. Hence $(EM)\mathbf{v} = E(M\mathbf{v}) = E\mathbf{0} = \mathbf{0}$, so $\mathbf{v}$ is in null space of EM. [This establishes that if $\mathbf{v}$ is in the null space of M then $\mathbf{v}$ is in the null space of EM.] On the other hand, if $\mathbf{v}$ is in the null space of EM then $(EM)\mathbf{v} = \mathbf{0}$. Hence $E(M\mathbf{v}) = \mathbf{0}$ so $M\mathbf{v} = E^{-1}\mathbf{0} = \mathbf{0}$ and $\mathbf{v}$ is in the null space of M. [This establishes that if $\mathbf{v}$ is in the null space of EM then $\mathbf{v}$ is in the null space of M.] Altogether this means the null space of M and the null space of EM contain exactly the same elements. We have thus established that the following two statements are equivalent for any $m \times n$ matrix M.

1. $\mathbf{v}$ is in the null space of M.

2. $\mathbf{v}$ is in the null space of EM.

To take it a step further, this means a certain set of columns of M are linearly dependent if and only if the same set of columns of EM are linearly dependent.

Thinking of matrix-vector multiplication as taking a linear combination of the columns of the matrix, any linear combination of the columns of M can be expressed as $M\mathbf{v}$ for some vector $\mathbf{v}$. To say that some set of columns of M are linearly dependent is to say there is a nonzero vector $\mathbf{v}$ such that $M\mathbf{v} = \mathbf{0}$. The nonzero entries of $\mathbf{v}$ determine the set of columns. Since $M\mathbf{v} = \mathbf{0}$ if and only if $E(M\mathbf{v}) = E\mathbf{0}$ (because E is invertible) if and only if $(EM)\mathbf{v} = \mathbf{0}$, a certain set of columns of M is linearly independent if and only if the corresponding set of columns of EM is linearly independent. More precisely, the exact same linear combination of columns of M that sums to zero, taken of EM instead, will also sum to zero. **Row operations do not affect the linear dependence relationships among the columns of a matrix**.

Crumpet 23: Uniqueness of Reduced Row Echelon Form

The fact that row operations do not affect the linear dependence relationships among the columns of a matrix lies at the heart of a proof that *the reduced row echelon form of a matrix exists and is unique*. The row reduction algorithm provides existence.

Suppose that an $m \times n$ matrix M has two reduced row echelon forms, A and B. The pivot columns of A and B are exactly those columns that are linearly independent of the columns to their left. This follows from the facts that (i) each pivot column of a matrix in reduced row echelon form is linearly independent of the columns on its left (it has a nonzero entry in a row where all the columns to its left have zeros); and (ii) each non-pivot column is linearly dependent on the columns to its left (it can be written as a linear combination of those columns). Because row operations do not alter the linear dependence relations among the columns of a matrix (and A and B are the results of series of row operations on M), the pivot columns of A must be the same as those of B. Since the pivot columns of a reduced row echelon form contain a 1 in the pivot position and zeros elsewhere, the pivot columns of A and B are in fact equal.

Suppose $A_{:,j}$ is a nonpivot column of A. Then $A_{:,j}$ and the pivot columns to its left (if any) form a linearly dependent set. A nontrivial linear combination of them sums to $\mathbf{0}$. Since $B_{:,j}$ has the same linear dependence relation with the pivot columns to its left, the *same* (corresponding) linear combination of $B_{:,j}$ and the pivot columns to its left sums to $\mathbf{0}$. Since the pivot columns of A and B are equal, it follows that $B_{:,j} = A_{:,j}$. This completes the proof.

Bases for the null space and column space of a matrix

The very nature of the row reduction algorithm followed by writing solutions of the homogeneous equation $M\mathbf{v} = \mathbf{0}$ in parametric form provides a basis for the null space. Each free variable gives rise to one vector in the parametric vector form (see section 3.7). The collection of all the vectors from this form comprise a basis.

In full detail, if $v_{f_1}, v_{f_2}, \ldots, v_{f_k}$ are the free variables of the linear system $M\mathbf{v} = \mathbf{0}$, then the row reduction algorithm leads to a solution of the (parametric vector) form

$$\mathbf{v} = r_1 \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \end{bmatrix} + r_2 \begin{bmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{n,2} \end{bmatrix} + \cdots + r_k \begin{bmatrix} a_{1,k} \\ a_{2,k} \\ \vdots \\ a_{n,k} \end{bmatrix}.$$

This form constitutes all the solutions of the homogeneous equation, so the columns of the matrix

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,k} \end{bmatrix}$$

span the null space of M. The algorithm also provides that $a_{f_1,1} = 1$ while $a_{f_1,2} = a_{f_1,3} = \cdots = a_{f_1,k} = 0$. Similarly, the entries of $A_{f_2,:}$ are all zero except the second; the entries of row $A_{f_3,:}$ are all zero except the third; and so on. Consequently the columns of A are linearly independent. Hence the columns of A (the vectors of the parametric form of the solution) are a linearly independent spanning set—a basis—for the null space of M.

Now suppose $v_{b_1}, v_{b_2}, \ldots, v_{b_\ell}$ are the basic variables for the linear system $M\mathbf{v} = \mathbf{b}$ and $b_1 < b_2 < \cdots < b_\ell$. We will argue that columns $M_{:,b_1}, M_{:,b_2}, \ldots, M_{:,b_\ell}$ form a basis (linearly independent spanning set) for the column space C. To see that these columns are linearly independent, let R be the reduced row echelon form of M. Then $R_{:,b_j}$ (column b_j of R) cannot be written as a linear combination of columns $R_{:,b_1}, R_{:,b_2}, \ldots, R_{:,b_{j-1}}$ (the columns to the left of column b_j corresponding to basic variables) for any $j > 1$. This is clear since $R_{j,b_j} = 1$ while $R_{j,b_1} = R_{j,b_2} = \cdots = R_{j,b_{j-1}} = 0$. This is sufficient to conclude that $R_{:,b_1}, R_{:,b_2}, \ldots, R_{:,b_\ell}$ are linearly independent. Can you show that, in general, the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, $\mathbf{v}_1 \neq \mathbf{0}$ are linearly dependent if and only if there is a $k > 1$ such that $\mathbf{v}_k$ can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}$. Answer on page 165. This completes the argument that $R_{:,b_1}, R_{:,b_2}, \ldots, R_{:,b_\ell}$ are linearly independent. Because row operations do not affect the linear dependence relationships among the columns of a matrix, we have that $M_{:,b_1}, M_{:,b_2}, \ldots, M_{:,b_\ell}$ are linearly independent.

To see that $M_{:,b_1}, M_{:,b_2}, \ldots, M_{:,b_\ell}$ span the column space of M we will rely on the fact that if $\mathbf{v}$ is in the span of (the arbitrary set of vectors) $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, then $\operatorname{span} V = \operatorname{span}(V \cup \{\mathbf{v}\})$. Can you support this claim? Answer on page 165. To see why this is useful, note that if column f of R corresponds to a free variable, then $R_{:,f}$ is a linear combination of the columns of basic variables to the left of $R_{:,f}$, say $R_{:,b_1}, R_{:,b_2}, \ldots, R_{:,b_j}$ where $b_j < f < b_{j+1}$. This is because R_{b_1}, the leftmost column corresponding to a basic variable, has a 1 in its first entry and zeros elsewhere; R_{b_2} has a 1 in its second entry and zeros elsewhere; R_{b_3} has a 1 in its third entry and zeros elsewhere; and so on. By construction, $R_{:,f}$ cannot have a nonzero entry below row b_j (if it did, it would contain a leading entry and therefore not be the column of a free variable). In symbols,

$$\begin{bmatrix} R_{:,b_1} & R_{:,b_2} & \cdots & R_{:,b_j} & R_{:,f} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & R_{1,f} \\ 0 & 1 & \cdots & 0 & R_{2,f} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & R_{j,f} \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Hence, $R_{:,f}$ is a linear combination of $R_{:,b_1}, R_{:,b_2}, \ldots, R_{:,b_\ell}$. Because row operations do not affect the linear dependence relationships among the columns of a matrix, we have that $M_{:,f}$ is a linear combination of $M_{:,b_1}, M_{:,b_2}, \ldots, M_{:,b_\ell}$. In other words, $M_{:,f}$ is in the span of $M_{:,b_1}, M_{:,b_2}, \ldots, M_{:,b_\ell}$. Repeatedly adding the columns of free variables (which are in the span of the columns of basic variables) to the set $\{M_{:,b_1}, M_{:,b_2}, \ldots, M_{:,b_\ell}\}$ leads to the conclusion that

$$\begin{aligned} \operatorname{span}\{M_{:,b_1}, M_{:,b_2}, \ldots, M_{:,b_\ell}\} &= \operatorname{span}\{M_{:,1}, M_{:,2}, \ldots, M_{:,n}\} \\ &= \text{column space of } M. \end{aligned}$$

Hence $\{M_{:,b_1}, M_{:,b_2}, \ldots, M_{:,b_\ell}\}$ is a linearly independent spanning set—a basis—for the column space of M

The preceding discussion justifies two general statements about any $m \times n$ matrix M, the combination of which leads to one of the most fundamental theorems of linear agebra.

- a basis for the null space of M can be formed using one vector for each free variable of the linear system $M\mathbf{v} = \mathbf{b}$.
- a basis for the column space of M is formed from the columns of M corresponding to the basic variables of the linear system $M\mathbf{v} = \mathbf{b}$.

These statements mean the rank of M equals the number of basic variables of $M\mathbf{v} = \mathbf{b}$, and the nullity of M equals the number of free variables of M. Since $M\mathbf{v} = \mathbf{b}$ has n variables in total, we have the following theorem.

Theorem 12. **[*Rank* and Nullity]** *If M is an $m \times n$ matrix, then the rank of M and the nullity of M sum to n.*

General Solutions

Let $\mathbf{v}_p$ be a particular (any single) solution of $M\mathbf{v} = \mathbf{b}$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be a basis for the null space of M. Now suppose $\mathbf{v}$ is any solution of $M\mathbf{v} = \mathbf{b}$. Then

$$\begin{aligned} M(\mathbf{v} - \mathbf{v}_p) &= M\mathbf{v} - M\mathbf{v}_p \\ &= \mathbf{b} - \mathbf{b} \\ &= \mathbf{0}. \end{aligned}$$

By definition, $\mathbf{v} - \mathbf{v}_p$ is in the null space of M. Because $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is a basis for the null space of M, there are coefficients $a_1, a_2, \ldots, a_k$ such that $\mathbf{v} - \mathbf{v}_p = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k$. Hence

$$\mathbf{v} = \mathbf{v}_p + a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k.$$

On the other hand, if $\mathbf{v} = \mathbf{v}_p + a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k$ for some coefficients $a_1, a_2, \ldots, a_k$ and particular solution $\mathbf{v}_p$, then

$$\begin{aligned} M\mathbf{v} &= M\left(\mathbf{v}_p + a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k\right) \\ &= M\mathbf{v}_p + M\left(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k\right) \\ &= \mathbf{b} + \mathbf{0} \\ &= \mathbf{b}. \end{aligned}$$

To summarize, these comments justify the following theorem.

Theorem 13. **[Characterization of Solutions of a Linear System]** *For a consistent linear system $M\mathbf{v} = \mathbf{b}$, the solution set is*

$$\mathbf{v} = \mathbf{v}_p + a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k$$

where $\mathbf{v}_p$ is any particular solution of $M\mathbf{v} = \mathbf{b}$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is a basis for the null space of M.

Key Concepts

column space of an $m \times n$ matrix M is $\text{span}\{M_{:,1}, M_{:,2}, \ldots, M_{:,n}\} = \{M\mathbf{v} : \mathbf{v} \in \mathbb{R}^n\}$.

null space of an $m \times n$ matrix M is $\{\mathbf{v} \in \mathbb{R}^n : M\mathbf{v} = \mathbf{0}\}$, the solution set of $M\mathbf{v} = \mathbf{0}$.

eigenspace of M corresponding to λ is the null space of $M - \lambda I$.

vector spaces the column space of a matrix and the null space of a matrix are vector spaces.

column space basis the columns of M corresponding to basic variables form a basis for the column space of M.

null space basis the vectors in the parametric vector form of the solution of $M\mathbf{v} = \mathbf{0}$ form a basis for the null space of M.

characterization of solutions of a linear system see theorem 13.

row operations (i) do not affect the null space of a matrix, and (ii) do not affect the linear dependence relationships of the columns of a matrix.

rank the dimension of the column space of a matrix.

nullity the dimension of the null space of a matrix.

Exercises

1. Is $\mathbf{b}$ in the column space of M?

   (a) $M = \begin{bmatrix} -1 & 4 \\ -4 & 15 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 3 \\ 10 \end{bmatrix}$

   (b) $M = \begin{bmatrix} 14 & -12 \\ 7 & -6 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

   (c) $M = \begin{bmatrix} -40 & -60 \\ 32 & 48 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 15 \\ -12 \end{bmatrix}$

   (d) $M = \begin{bmatrix} 27 & 33 \\ 9 & 11 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$ [$\mathbb{S}$]-315

   (e) $M = \begin{bmatrix} 2 & -6 \\ 4 & -14 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$ [$\mathbb{A}$]-355

   (f) $M = \begin{bmatrix} 10 & 6 \\ -40 & -24 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ [$\mathbb{A}$]-355

2. Find a basis for the column space of M from question 1. [$\mathbb{S}$]-315 [$\mathbb{A}$]-355

3. Find a basis for the null space of M from question 1. [$\mathbb{S}$]-315 [$\mathbb{A}$]-355

4. R is the reduced row echelon form of $\begin{bmatrix} M & \mathbf{b} \end{bmatrix}$, the matrix M augmented by some vector $\mathbf{b}$. (i) Is $\mathbf{b}$ in the column space of M? (ii) Find a basis for the column space of M. (iii) Find a basis for the null space of M.

   (a) $M = \begin{bmatrix} -20 & -90 & 6 & 153 \\ -6 & -27 & 2 & 45 \\ 4 & 18 & -2 & -36 \end{bmatrix}$

   $R = \begin{bmatrix} 1 & 9/2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 6 \\ 0 & 0 & 0 & 1 & -10/9 \end{bmatrix}$ [$\mathbb{S}$]-315

   (b) $M = \begin{bmatrix} -63 & 14 & 77 & 3 \\ -36 & 8 & 44 & 1 \\ 54 & -12 & -66 & -2 \end{bmatrix}$

   $R = \begin{bmatrix} 1 & -2/9 & -11/9 & 0 & 7/9 \\ 0 & 0 & 0 & 1 & -9 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

   (c) $M = \begin{bmatrix} -154 & -30 & -9 & -76 & 35 \\ 121 & 20 & 9 & 56 & -28 \\ -33 & -5 & -3 & -14 & 7 \end{bmatrix}$

   $R = \begin{bmatrix} 1 & 0 & 0 & 8/11 & -7/11 & 1 \\ 0 & 1 & 0 & -2/5 & 7/5 & 2 \\ 0 & 0 & 1 & -8/3 & 7/3 & 1 \end{bmatrix}$

   (d) $M = \begin{bmatrix} 187 & 99 & -74 & -12 \\ 0 & -11 & -5 & 4 \\ -33 & -22 & 11 & 4 \\ -154 & -77 & 63 & 12 \end{bmatrix}$

   $R = \begin{bmatrix} 1 & 0 & -7/11 & 0 & 0 \\ 0 & 1 & 5/11 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ [$\mathbb{S}$]-316

   (e) $M = \begin{bmatrix} 27 & 0 & -3 & 23 \\ -126 & 88 & 9 & -103 \\ 45 & -33 & -3 & 37 \\ -117 & 77 & 6 & -102 \end{bmatrix}$

   $R = \begin{bmatrix} 1 & 0 & 0 & 10/9 & 0 \\ 0 & 1 & 0 & 2/11 & 0 \\ 0 & 0 & 1 & 7/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

   (f) $M = \begin{bmatrix} -12 & 6 & 22 & 11 \\ -36 & 15 & 58 & 33 \\ 24 & -9 & -36 & -11 \\ 60 & -30 & -110 & -33 \end{bmatrix}$

   $R = \begin{bmatrix} 1 & 0 & -1/2 & 0 & -1 \\ 0 & 1 & 8/3 & 0 & -1/3 \\ 0 & 0 & 0 & 1 & -9/11 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ [$\mathbb{A}$]-355

   (g) $M = \begin{bmatrix} -12 & 32 & -36 & 8 \\ 9 & -24 & 27 & -6 \\ -30 & 80 & -90 & 20 \\ -3 & 8 & -9 & 2 \\ 24 & -64 & 72 & -16 \end{bmatrix}$

   $R = \begin{bmatrix} 1 & -8/3 & 3 & -2/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

   (h) $M = \begin{bmatrix} -12 & -8 & 8 & -48 \\ -36 & -20 & 16 & -132 \\ 24 & 24 & 24 & 156 \\ -24 & -8 & -16 & -84 \\ 36 & 32 & -72 & 132 \end{bmatrix}$

   $R = \begin{bmatrix} 1 & 0 & 0 & 0 & 2/3 \\ 0 & 1 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & 3/2 \\ 0 & 0 & 0 & 1 & 11/12 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ [$\mathbb{A}$]-355

   (i) $M = \begin{bmatrix} -108 & 54 & -63 & 40 \\ -120 & 60 & -70 & 5 \\ -60 & 30 & -35 & 50 \\ -24 & 12 & -14 & -10 \\ -48 & 24 & -28 & 5 \end{bmatrix}$

   $R = \begin{bmatrix} 1 & -1/2 & 7/12 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

(j) $M = \begin{bmatrix} -5 & -14 & -3 & 20 & -22 \\ 3 & 7 & 0 & -20 & 22 \\ -13 & -21 & -24 & 90 & -77 \\ 3 & 7 & -3 & -30 & 33 \\ 10 & 21 & 6 & -80 & 77 \end{bmatrix}$

$R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -3 \\ 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 & 14 \\ 0 & 0 & 0 & 1 & 0 & -8 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{bmatrix}$ [𝔸]-355

5. Make general statements about how the reduced row echelon form of $\begin{bmatrix} M & \mathbf{b} \end{bmatrix}$ helps (i) determine whether $\mathbf{b}$ is in the column space of M; (ii) find a basis for the column space of M; and (iii) find a basis for the null space of M.

6. M row reduces to R. (i) Verify that $\mathbf{b}$ is in the column space of M and (ii) find the general solution of $M\mathbf{v} = \mathbf{b}$.

(a) $M = \begin{bmatrix} -15 & -35 & -40 & -7 & -27 \\ 27 & 63 & 72 & 42 & 15 \\ 9 & 21 & 24 & 7 & 13 \end{bmatrix}$

$R = \begin{bmatrix} 1 & 7/3 & 8/3 & 0 & 7/3 \\ 0 & 0 & 0 & 1 & -8/7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

$\mathbf{b} = \begin{bmatrix} -22 \\ 69 \\ 16 \end{bmatrix}$ [𝕊]-316

(b) $M = \begin{bmatrix} 14 & -49 & 3 & 54 & -35 \\ 4 & -14 & 3 & 24 & -15 \\ -4 & 14 & -6 & -36 & 22 \end{bmatrix}$

$R = \begin{bmatrix} 1 & -7/2 & 0 & 3 & -2 \\ 0 & 0 & 1 & 4 & -7/3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

$\mathbf{b} = \begin{bmatrix} 11 \\ 1 \\ 2 \end{bmatrix}$

(c) $M = \begin{bmatrix} -91 & 104 & -2 & -18 \\ 56 & -64 & 2 & 9 \\ -154 & 176 & -6 & -24 \end{bmatrix}$

$R = \begin{bmatrix} 1 & -8/7 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

$\mathbf{b} = \begin{bmatrix} -8 \\ -1 \\ 6 \end{bmatrix}$

(d) $M = \begin{bmatrix} 3 & -21 & -12 & 18 \\ 9 & -56 & -32 & 50 \\ -6 & 35 & 20 & -30 \end{bmatrix}$

$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 4/7 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

$\mathbf{b} = \begin{bmatrix} 3 \\ 8 \\ -5 \end{bmatrix}$ [𝔸]-355

(e) $M = \begin{bmatrix} 21 & -35 & 15 & -84 \\ 6 & -10 & 5 & -25 \\ -18 & 30 & -10 & 68 \end{bmatrix}$

$R = \begin{bmatrix} 1 & -5/3 & 0 & -3 \\ 0 & 0 & 1 & -7/5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

$\mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 2 \end{bmatrix}$

7. Given an arbitrary $m \times n$ matrix M and a vector $\mathbf{b} \neq \mathbf{0}$, is the solution set of $M\mathbf{v} = \mathbf{b}$ a vector space? Explain.

8. Columns 2 and 5 of matrix M form a basis for the column space of M. Use this information to help decide whether $M\mathbf{v} = \mathbf{b}$ is consistent.
$M = \begin{bmatrix} 90 & 240 & -120 & 240 & -40 \\ 3 & 8 & -4 & 8 & 0 \\ -48 & -128 & 64 & -128 & 20 \\ -108 & -288 & 144 & -288 & 45 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 200 \\ 16 \\ -116 \\ -261 \end{bmatrix}$

9. Let M be a 4×5 coefficient matrix with columns 2,3, and 4 representing free variables. Argue that the set containing the fifth column of M and any one of the first four form a basis for the column space of M.

10. Let $M = \begin{bmatrix} 3 & 10 & 4 & 7 \\ 9 & 45 & 18 & 7 \\ 9 & 25 & 10 & 28 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 10 \\ 45 \\ 25 \end{bmatrix}$.

    (a) Solve $M\mathbf{v} = \mathbf{b}$ by inspection. (Row reduce or use SageMath if you don't see it, but then reflect on why you did not see it.)

    (b) Use the fact that $\left\{ \begin{bmatrix} 0 \\ -2 \\ 5 \\ 0 \end{bmatrix} \right\}$ is a basis for the null space of M to write down all the solutions of $M\mathbf{v} = \mathbf{b}$ in parametric vector form.

    (c) Write down three distinct solutions of $M\mathbf{v} = \mathbf{b}$ all different from the solution in (a).

11. Let $M = \begin{bmatrix} 4 & 5 & -5 & -12 \\ 2 & 5 & -5 & -3 \\ 0 & -15 & 15 & -15 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 24 \\ 6 \\ 30 \end{bmatrix}$. [𝕊]-317

    (a) Find one solution of $M\mathbf{v} = \mathbf{b}$ by inspection. (Row reduce or use SageMath if you don't see it, but then reflect on why you did not see it.)

    (b) Use the fact that $\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\}$ is a basis for the null space of M to help write down all the solutions of $M\mathbf{v} = \mathbf{b}$ in parametric vector form.

    (c) Write down three distinct solutions of $M\mathbf{v} = \mathbf{b}$ all different from the solution in (a).

12. Let $M = \begin{bmatrix} 154 & 242 & 15 & -9 \\ 63 & 99 & 5 & -3 \\ -112 & -176 & -10 & 6 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 396 \\ 162 \\ -288 \end{bmatrix}$.
    (a) Solve $M\mathbf{v} = \mathbf{b}$ by inspection. (Row reduce or use SageMath if you don't see it, but then reflect on why you did not see it.)
    (b) Use the fact that $\left\{ \begin{bmatrix} -11 \\ 7 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 3 \\ 5 \end{bmatrix} \right\}$ is a basis for the null space of M to write down all the solutions of $M\mathbf{v} = \mathbf{b}$ in parametric vector form.
    (c) Write down three distinct solutions of $M\mathbf{v} = \mathbf{b}$ all different from the solution in (a).

13. Use the fact that row operations were used to reduce $A = \begin{bmatrix} -16 & 72 & -15 & -91 & -182 & -6 \\ 6 & -27 & 5 & 33 & 63 & 3 \\ 2 & -9 & 5 & 17 & 14 & 6 \\ -10 & 45 & -5 & -49 & -91 & -6 \end{bmatrix}$ to $B = \begin{bmatrix} 2 & -9 & 0 & 8 & 0 & 0 \\ 0 & 0 & 5 & 9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}$ to find a set of three columns of A that are linearly independent. Are there other such sets? [𝔸]-355

14. Let $A = \begin{bmatrix} 8 & -16 & -18 \\ 16 & k & -45 \\ 8 & -32 & -36 \end{bmatrix}$, $B = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 8 & 9 \\ 0 & 0 & 0 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 0 \\ 9 \\ -8 \end{bmatrix}$.
    (a) Verify that $B\mathbf{v} = \mathbf{0}$.
    (b) Given that A row reduces to B, find k.

15. What is the dimension of the eigenspace corresponding to (the eigenvalue) λ?
    (a) $\begin{bmatrix} 3 & 15 & -12 \\ 3 & 7 & 4 \\ 6 & -10 & 20 \end{bmatrix}$; $\lambda = 6$
    (b) $\begin{bmatrix} 47 & 45 & -75 \\ 5 & 31 & -15 \\ 15 & 27 & -23 \end{bmatrix}$; $\lambda = 22$ [𝕊]-317
    (c) $\begin{bmatrix} 18 & 6 & -4 \\ -10 & -1 & 10 \\ -4 & -6 & 18 \end{bmatrix}$; $\lambda = 14$
    (d) $\begin{bmatrix} 6 & -20 & 12 \\ 16 & 42 & -12 \\ 20 & 25 & 7 \end{bmatrix}$; $\lambda = 11$ [𝔸]-355

16. Find a basis for the eigenspace corresponding to (the eigenvalue) λ in question 15.

17. What is the rank of a matrix of all zeros?

18. What is the rank of a matrix of all ones? [𝔸]-355

19. What is the rank of a matrix of all twos?

20. Given an $m \times n$ matrix M and invertible $n \times n$ matrix P, show that the rank of MP equals the rank of M by the following argument. Let C_M be the column space of M and C_{MP} be the column space of MP.
    (a) Suppose $\mathbf{v}$ is in C_M, and show that $\mathbf{v}$ is in C_{MP}.
    (b) Suppose $\mathbf{v}$ is in C_{MP}, and show that $\mathbf{v}$ is in C_M.
    (c) Conclude that the rank of MP equals the rank of M.

21. Given an $m \times n$ matrix M and invertible $m \times m$ matrix Q, show that the nullity of QM equals the nullity of M by the following argument. Let N_M be the null space of M and N_{QM} be the null space of QM.
    (a) Suppose $\mathbf{v}$ is in N_M, and show that $\mathbf{v}$ is in N_{QM}.
    (b) Suppose $\mathbf{v}$ is in N_{QM}, and show that $\mathbf{v}$ is in N_M.
    (c) Conclude that the nullity of QM equals the nullity of M.

Answers

linear dependence Show that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, $\mathbf{v}_1 \neq \mathbf{0}$ are linearly dependent if and only if there is a $k > 1$ such that $\mathbf{v}_k$ can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}$. Supposing $\mathbf{v}_k$ can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}$, we have immediately that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are linearly dependent. Now suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, $\mathbf{v}_1 \neq \mathbf{0}$ are linearly dependent. Then there exists a linear combination

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_p\mathbf{v}_p = \mathbf{0}.$$

It must be that at least one of $a_2, a_3, \ldots, a_p$ is nonzero since $\mathbf{v}_1 \neq \mathbf{0}$. Set $k = \max\{i : a_i \neq 0\}$. Then $k > 1$, $a_k \neq 0$, and $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k = \mathbf{0}$, so

$$\mathbf{v}_k = -\frac{a_1}{a_k}\mathbf{v}_1 - \frac{a_2}{a_k}\mathbf{v}_2 - \cdots - \frac{a_{k-1}}{a_k}\mathbf{v}_{k-1}.$$

span Show that if $\mathbf{v}$ is in the span of (the arbitrary set of vectors) $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, then $\operatorname{span} V = \operatorname{span}(V \cup \{\mathbf{v}\})$. First, $\operatorname{span} V \subseteq \operatorname{span}(V \cup \{\mathbf{v}\})$ always since every linear combination of the vectors in V is also a linear combination of vectors in $V \cup \{\mathbf{v}\}$ (with the coefficient of $\mathbf{v}$ equal 0, for example). It remains to show that

$\operatorname{span}(V \cup \{\mathbf{v}\}) \subseteq \operatorname{span} V$. In other words, if $\mathbf{w} \in \operatorname{span}(V \cup \{\mathbf{v}\})$ then $\mathbf{w} \in \operatorname{span} V$. To that end, suppose $\mathbf{w} \in \operatorname{span}(V \cup \{\mathbf{v}\})$ and (a) write $\mathbf{w}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p, \mathbf{v}$; and (b) write $\mathbf{v}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ (which is possible since $\mathbf{v}$ is in the span of V):

$$\begin{aligned} \mathbf{w} &= a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_p\mathbf{v}_p + a\mathbf{v} \\ \mathbf{v} &= b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_p\mathbf{v}_p. \end{aligned}$$

By substitution,

$$\begin{aligned} \mathbf{w} &= a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_p\mathbf{v}_p + a(b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_p\mathbf{v}_p) \\ &= (a_1 + ab_1)\mathbf{v}_1 + (a_2 + ab_2)\mathbf{v}_2 + \cdots + (a_p + ab_p)\mathbf{v}_p \end{aligned}$$

and therefore $\mathbf{w} \in \operatorname{span} V$.

5.2 Coordinate Vectors [4.1, 4.2]

In section 4.2 it was noted that given a basis for a vector space, each vector has a unique (exactly one) representation as a linear combination of the vectors in the basis. That means there is a one-to-one correspondence between elements in the vector space and linear combinations of vectors in the basis. Each vector in the vector space can be identified by its corresponding linear combination, or more succinctly, by the list of coefficients in that linear combination. These coefficient lists serve as unique identifiers for the vectors much the same way social security numbers serve as unique identifiers for people. A country with a social security system assigns each of its citizens exactly one social security number. Each citizen has one social security number and each social security number identifies one person.

The situation is a little more complicated when a person is a citizen of more than one country with a social security system. The same person will have multiple social security numbers, one for each country of which they are a citizen. One social security number may be useful in France while another is useful in Mali. In a similar way, a single vector will have multiple unique identifiers, one for each basis of the vector space. Each basis gives a different labeling system for the vectors (citizens) of its vector space.

As pointed out in section 4.2,

$$\mathcal{E} = \{I_{:,1}, I_{:,2}, \dots, I_{:,n}\}$$

is a basis for the vector space $\mathcal{M}_{n\times 1}(\mathbb{R})$ (the collection of all $n \times 1$ column matrices with real coefficients), better known as $\mathbb{R}^n$[1]. Writing a vector in $\mathbb{R}^n$ as a linear combination of these basis elements is a simple matter:

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T = x_1 I_{:,1} + x_2 I_{:,2} + \cdots + x_n I_{:,n},$$

and this is the only such linear combination. If we write the coefficients as a column vector decorated with the name of the basis as in $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{\mathcal{E}}$, we have what is known as a **coordinate vector**. The vectors we have been writing all along have been written with respect to the standard basis, and we will maintain this practice. A vector written with no basis indicated means that it has been written with respect to the standard basis. Hence $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{\mathcal{E}}$ and $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ represent the same vector. Given a different basis, say $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, writing $\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$ as a linear combination may require different coefficients:

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

for some $c_1, c_2, \dots, c_n$. The coefficients of this linear combination, in order, form the unique identifier of $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ in the context of the basis $\mathcal{V}$: $\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}_{\mathcal{V}}$, read as the coordinate vector with respect to $\mathcal{V}$.

It should be noted that the order of the elements in a basis matters, and that the entries in the coordinate vector must correspond to this ordering. The first entry of the coordinate vector corresponds with the first vector in the basis. The second entry of the coordinate vector corresponds with the second vector in the basis. And so on. This correspondence is required to maintain the uniqueness of representation. Different orderings of the same set of vectors provide different coordinate systems for the vector space. A basis is thus an ordered set.

[1]Technically, $\mathcal{M}_{n\times 1}(\mathbb{R})$ and $\mathbb{R}^n$ are isomorphic (see section 4.5)

The basis $\mathcal{B} = \{I_{:,2}, I_{:,1}, \ldots, I_{:,n}\}$ is different from the standard basis $\mathcal{E}$ (even though, as unordered sets, $\mathcal{B}$ and $\mathcal{E}$ are equal). Of course, it is still true that

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T = x_1 I_{:,1} + x_2 I_{:,2} + \cdots + x_n I_{:,n}.$$

This fact will never change. But this means the coordinate vector of $\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$ with respect to $\mathcal{B}$ is $\begin{bmatrix} x_2 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}_{\mathcal{B}}$. Hence

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{\mathcal{E}} = \begin{bmatrix} x_2 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}_{\mathcal{B}}.$$

Letting $C = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$, can you verify that

$$\begin{bmatrix} 4 \\ 6 \\ -3 \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} 6 \\ 4 \\ -3 \end{bmatrix} = \begin{bmatrix} 2 \\ 7 \\ -3 \end{bmatrix}_{C}?$$

Answer on page 174. As a matter of notation, when the basis with respect to which a vector $\mathbf{v}$ is written is important, we will enclose the name of the vector in square brackets and subscript it with the name of the basis as in $[\mathbf{v}]_{\mathcal{B}}$.

Notice that

$$\begin{bmatrix} 2 \\ 7 \\ -3 \end{bmatrix}_{C} = 2\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 7\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - 3\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 7 \\ -3 \end{bmatrix}$$

and

$$\begin{bmatrix} 4 \\ 6 \\ -3 \end{bmatrix}_{\mathcal{B}} = 4\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 6\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - 3\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 4 \\ 6 \\ -3 \end{bmatrix}.$$

In general, if $\mathcal{B}$ is a basis of $\mathbb{R}^n$ and $[\mathcal{B}]_{\mathcal{E}}$ is the matrix whose columns are the vectors of $\mathcal{B}$ written with respect to the standard basis respecting order, then

$$\mathbf{v} = [\mathbf{v}]_{\mathcal{E}} = [\mathcal{B}]_{\mathcal{E}}\,[\mathbf{v}]_{\mathcal{B}} \tag{5.2.1}$$

Note, however, there is nothing special about the standard basis beyond the fact that it is the most familiar. If C is a basis of $\mathbb{R}^n$ and $[\mathcal{B}]_C$ is the matrix whose columns are the vectors of $\mathcal{B}$ written with respect to the basis C respecting order, then

$$\mathbf{v} = [\mathbf{v}]_C = [\mathcal{B}]_C\,[\mathbf{v}]_{\mathcal{B}}\,. \tag{5.2.2}$$

This can be verified by direct calculation. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ and $C = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ and write the vectors of $\mathcal{B}$ with respect to C:

$$\begin{aligned} \mathbf{b}_1 &= M_{1,1}\mathbf{c}_1 + M_{2,1}\mathbf{c}_2 + \cdots + M_{n,1}\mathbf{c}_n \\ \mathbf{b}_2 &= M_{1,2}\mathbf{c}_1 + M_{2,2}\mathbf{c}_2 + \cdots + M_{n,2}\mathbf{c}_n \\ &\vdots \\ \mathbf{b}_n &= M_{1,n}\mathbf{c}_1 + M_{2,n}\mathbf{c}_2 + \cdots + M_{n,n}\mathbf{c}_n \end{aligned} \tag{5.2.3}$$

and $\mathbf{v}$ with respect to $\mathcal{B}$:

$$\mathbf{v} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n. \tag{5.2.4}$$

Then

$$[\mathcal{B}]_{C} = \begin{bmatrix} M_{1,1} & M_{1,2} & \cdots & M_{1,n} \\ M_{2,1} & M_{2,2} & \cdots & M_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{n,1} & M_{n,2} & \cdots & M_{n,n} \end{bmatrix}$$

and

$$\begin{aligned}
[\mathcal{B}]_{C}\,[\mathbf{v}]_{\mathcal{B}} &= \begin{bmatrix} M_{1,1} & M_{1,2} & \cdots & M_{1,n} \\ M_{2,1} & M_{2,2} & \cdots & M_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{n,1} & M_{n,2} & \cdots & M_{n,n} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v_1 \begin{bmatrix} M_{1,1} \\ M_{2,1} \\ \vdots \\ M_{n,1} \end{bmatrix} + v_2 \begin{bmatrix} M_{1,2} \\ M_{2,2} \\ \vdots \\ M_{n,2} \end{bmatrix} + \cdots + v_n \begin{bmatrix} M_{1,n} \\ M_{2,n} \\ \vdots \\ M_{n,n} \end{bmatrix} \\
&= \begin{bmatrix} v_1 M_{1,1} + v_2 M_{1,2} + \cdots + v_n M_{1,n} \\ v_1 M_{2,1} + v_2 M_{2,2} + \cdots + v_n M_{2,n} \\ \vdots \\ v_1 M_{n,1} + v_2 M_{n,2} + \cdots + v_n M_{n,n} \end{bmatrix}. \qquad (5.2.5)
\end{aligned}$$

On the other hand, direct substitution of (5.2.3) into (5.2.4) yields

$$\begin{aligned}
\mathbf{v} &= v_1\left(M_{1,1}\mathbf{c}_1 + M_{2,1}\mathbf{c}_2 + \cdots + M_{n,1}\mathbf{c}_n\right) \\
&\quad + v_2\left(M_{1,2}\mathbf{c}_1 + M_{2,2}\mathbf{c}_2 + \cdots + M_{n,2}\mathbf{c}_n\right) \\
&\quad \vdots \\
&\quad + v_n\left(M_{1,n}\mathbf{c}_1 + M_{2,n}\mathbf{c}_2 + \cdots + M_{n,n}\mathbf{c}_n\right) \\
&= \left(v_1 M_{1,1} + v_2 M_{1,2} + \cdots + v_n M_{1,n}\right)\mathbf{c}_1 \\
&\quad + \left(v_1 M_{2,1} + v_2 M_{2,2} + \cdots + v_n M_{2,n}\right)\mathbf{c}_2 \\
&\quad \vdots \\
&\quad + \left(v_1 M_{n,1} + v_2 M_{n,2} + \cdots + v_n M_{n,n}\right)\mathbf{c}_n
\end{aligned}$$

which verifies that (5.2.5) is $[\mathbf{v}]_C$. Equation (5.2.2) is one formula for a so-called **change of basis**. It gives a formula for changing the basis with respect to which $\mathbf{v}$ is written from $\mathcal{B}$ to C.

Being a basis of $\mathbb{R}^n$, $\mathcal{B}$ contains n linearly independent vectors with n entries each, so $[\mathcal{B}]_C$ is an $n \times n$ matrix with linearly independent columns, making it an invertible matrix (theorem 7). In particular, if both bases $\mathcal{B}$ and C are written with respect to the standard basis and $\mathbf{v}$ is an arbitrary vector, we have $\mathbf{v} = [\mathcal{B}]_{\mathcal{E}}\,[\mathbf{v}]_{\mathcal{B}}$ and $\mathbf{v} = [C]_{\mathcal{E}}\,[\mathbf{v}]_C$, so

$$[\mathcal{B}]_{\mathcal{E}}\,[\mathbf{v}]_{\mathcal{B}} = [C]_{\mathcal{E}}\,[\mathbf{v}]_C$$

and we can left-multiply both sides of this equation by $[C]_{\mathcal{E}}^{-1}$, yielding

$$[\mathbf{v}]_C = [C]_{\mathcal{E}}^{-1}\,[\mathcal{B}]_{\mathcal{E}}\,[\mathbf{v}]_{\mathcal{B}}\,. \qquad (5.2.6)$$

Comparing this equation to (5.2.2) it must be that $[\mathcal{B}]_C = [C]_{\mathcal{E}}^{-1}\,[\mathcal{B}]_{\mathcal{E}}$.

In retrospect, this should not be surprising. Multiplying $[\mathbf{v}]_{\mathcal{B}}$ by $[\mathcal{B}]_{\mathcal{E}}$ gives $[\mathbf{v}]_{\mathcal{E}}$ (see equation (5.2.1)). This same equation tells us that multiplying $[\mathbf{v}]_{\mathcal{E}}$ by $[C]_{\mathcal{E}}^{-1}$ gives $[\mathbf{v}]_C$. Diagrammatically

$$[\mathbf{v}]_{\mathcal{B}} \stackrel{\text{times } [\mathcal{B}]_{\mathcal{E}}}{\longrightarrow} [\mathbf{v}]_{\mathcal{E}} \stackrel{\text{times } [C]_{\mathcal{E}}^{-1}}{\longrightarrow} [\mathbf{v}]_C\,,$$

another way to understand (5.2.6).

Key Concepts

coordinate vector If $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ is a basis for a vector space and $\mathbf{v} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n$, then

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}_{\mathcal{B}}$$

is the coordinate vector of $\mathbf{v}$ with respect to $\mathcal{B}$ and may be denoted $[\mathbf{v}]_{\mathcal{B}}$.

change of basis given bases $\mathcal{B}$ and C of a vector space V and $\mathbf{v} \in V$,

$$[\mathbf{v}]_C = [\mathcal{B}]_C\,[\mathbf{v}]_{\mathcal{B}} = [C]_{\mathcal{E}}^{-1}\,[\mathcal{B}]_{\mathcal{E}}\,[\mathbf{v}]_{\mathcal{B}}\,.$$

Exercises

1. Let $\mathcal{B} = \left\{ \begin{bmatrix} 6 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\}$. What vector $\mathbf{v}$ is given by the coordinate vector?
   (a) $\begin{bmatrix} 8 \\ 1 \end{bmatrix}_{\mathcal{B}}$ (b) $\begin{bmatrix} 5 \\ 4 \end{bmatrix}_{\mathcal{B}}$ (c) $\begin{bmatrix} -2 \\ 2 \end{bmatrix}_{\mathcal{B}}$ (d) $\begin{bmatrix} 5 \\ 5 \end{bmatrix}_{\mathcal{B}}$

2. Let $\mathcal{B} = \left\{ \begin{bmatrix} -2 \\ -4 \\ -2 \end{bmatrix}, \begin{bmatrix} 6 \\ 14 \\ 8 \end{bmatrix}, \begin{bmatrix} 9 \\ 15 \\ 9 \end{bmatrix} \right\}$. Find the vector $\mathbf{x}$ determined by the coordinate vector.
   (a) $\begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}_{\mathcal{B}}$ (b) $\begin{bmatrix} -4 \\ 2 \\ 0 \end{bmatrix}_{\mathcal{B}}$ (c) $\begin{bmatrix} 5 \\ -2 \\ 3 \end{bmatrix}_{\mathcal{B}}$ (d) $\begin{bmatrix} 3 \\ 5 \\ 4 \end{bmatrix}_{\mathcal{B}}$

3. Let $\mathcal{B} = \{-5x + 3,\ 3x + 8\}$. Find the vector $\mathbf{x}$ determined by the coordinate vector.
   (a) $\begin{bmatrix} -4 \\ -2 \end{bmatrix}_{\mathcal{B}}$ (b) $\begin{bmatrix} -1 \\ -1 \end{bmatrix}_{\mathcal{B}}$ (c) $\begin{bmatrix} 1 \\ 8 \end{bmatrix}_{\mathcal{B}}$ (d) $\begin{bmatrix} 5 \\ 0 \end{bmatrix}_{\mathcal{B}}$

4. Let $\mathcal{B} = \left\{ \begin{bmatrix} 9 & 2 \\ 5 & -3 \end{bmatrix}, \begin{bmatrix} -4 & 5 \\ 5 & 5 \end{bmatrix}, \begin{bmatrix} 3 & -1 \\ -2 & -2 \end{bmatrix}, \begin{bmatrix} 7 & 6 \\ 3 & -2 \end{bmatrix} \right\}$ and suppose $[\mathbf{x}]_{\mathcal{B}}$ is as given. Compute $\mathbf{x}$.
   (a) $\begin{bmatrix} 4 \\ -3 \\ 6 \\ 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 \\ 8 \\ 7 \\ -1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 \\ -5 \\ 1 \\ -5 \end{bmatrix}$ (d) $\begin{bmatrix} -3 \\ 3 \\ 8 \\ 5 \end{bmatrix}$

5. Write the vector $\mathbf{v} = 3 - 4t + 5t^2$ as a coordinate vector with respect to the basis $\mathcal{B}$. [$\mathcal{B}$ is linearly independent and therefore a basis for its span, and $\mathbf{v}$ is in span$\mathcal{B}$.]
   (a) $\mathcal{B} = \left\{3 - 4t, t^2\right\}$ [𝕊]-317
   (b) $\mathcal{B} = \left\{1, t, t^2\right\}$ [𝔸]-355
   (c) $\mathcal{B} = \left\{t^2, t, 1\right\}$
   (d) $\mathcal{B} = \left\{1, t, t^2, t^3\right\}$
   (e) $\mathcal{B} = \left\{3, -4t, 5t^2\right\}$
   (f) $\mathcal{B} = \left\{3 - 4t + 5t^2, 1 - t + 8t^2, 14 - 5t + 6t^2\right\}$

6. Write the vector $\mathbf{v} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ as a coordinate vector with respect to the basis $\mathcal{B}$. [$\mathcal{B}$ is linearly independent and therefore a basis for its span, and $\mathbf{v}$ is in span$\mathcal{B}$.]
   (a) $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix} \right\}$
   (b) $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 9 & 12 \end{bmatrix} \right\}$
   (c) $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$
   [𝕊]-317
   (d) $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$
   [𝔸]-355
   (e) $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$
   (f) $\mathcal{B} = \left\{ \begin{bmatrix} 3 & 8 \\ -2 & 9 \end{bmatrix}, \begin{bmatrix} 2 & 6 \\ -5 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} -3 & 7 \\ 5 & 4 \end{bmatrix} \right\}$

7. Write the vector $\mathbf{v} = \langle 6, 3, -4 \rangle$ as a coordinate vector with respect to the basis $\mathcal{B}$. [$\mathcal{B}$ is linearly independent and therefore a basis for its span, and $\mathbf{v}$ is in span$\mathcal{B}$.]
   (a) $\mathcal{B} = \{\langle 2, 1, 0\rangle, \langle 0, 0, 1\rangle\}$
   (b) $\mathcal{B} = \{\langle 1, 0, 0\rangle, \langle 0, 3, -4\rangle\}$
   (c) $\mathcal{B} = \{\langle 1, 0, 0\rangle, \langle 0, 1, 0\rangle, \langle 0, 0, 1\rangle\}$
   (d) $\mathcal{B} = \{\langle 6, 0, 0\rangle, \langle 0, 3, 0\rangle, \langle 0, 0, -4\rangle\}$
   (e) $\mathcal{B} = \{\langle 2, -1, 9\rangle, \langle 6, 3, -4\rangle, \langle -8, 1, 1\rangle\}$
   [𝕊]-317
   (f) $\mathcal{B} = \left\{\left\langle \frac{1}{2}, \frac{1}{4}, -\frac{1}{3} \right\rangle\right\}$ [𝔸]-355

8. Write the vector $\mathbf{v}$ as a coordinate vector with respect to the basis $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 5 \\ -9 \end{bmatrix} \right\}$.
   [𝕊]-318
   (a) $\mathbf{v} = \begin{bmatrix} 7 \\ -12 \end{bmatrix}$
   (b) $\mathbf{v} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$

9. Write the vector $\mathbf{v}$ as a coordinate vector with respect to the basis $\mathcal{B} = \left\{ \begin{bmatrix} 6 \\ 3 \end{bmatrix}, \begin{bmatrix} -5 \\ -8 \end{bmatrix} \right\}$ of $\mathbb{R}^2$. [𝔸]-355
   (a) $\mathbf{v} = \begin{bmatrix} 17 \\ 14 \end{bmatrix}$
   (b) $\mathbf{v} = \begin{bmatrix} -8 \\ 7 \end{bmatrix}$

10. Write the vector $\mathbf{v}$ as a coordinate vector with respect to the basis $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 9 \end{bmatrix}, \begin{bmatrix} -5 \\ 1 \end{bmatrix} \right\}$ of $\mathbb{R}^2$.
    (a) $\mathbf{v} = \begin{bmatrix} 17 \\ 15 \end{bmatrix}$
    (b) $\mathbf{v} = \begin{bmatrix} -3 \\ 19 \end{bmatrix}$

11. Write the vector $\mathbf{v}$ as a coordinate vector with respect to the basis $\mathcal{B} = \left\{ \begin{bmatrix} -4 \\ 7 \\ 3 \end{bmatrix}, \begin{bmatrix} 9 \\ 4 \\ 3 \end{bmatrix}, \begin{bmatrix} -3 \\ 5 \\ 4 \end{bmatrix} \right\}$ of $\mathbb{R}^3$. [$\mathbb{A}$]-355

    (a) $\mathbf{v} = \begin{bmatrix} 17 & 10 & 5 \end{bmatrix}^T$

    (b) $\mathbf{v} = \begin{bmatrix} 8 & 5 & 9 \end{bmatrix}^T$

    (c) $\mathbf{v} = \begin{bmatrix} 19 & 6 & 7 \end{bmatrix}^T$

12. Write the vector $\mathbf{v}$ as a coordinate vector with respect to the basis $\mathcal{B} = \left\{ \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 9 \end{bmatrix}, \begin{bmatrix} 6 \\ 4 \\ -9 \end{bmatrix} \right\}$ of $\mathbb{R}^3$.

    (a) $\mathbf{v} = \begin{bmatrix} 11 & 4 & 4 \end{bmatrix}^T$

    (b) $\mathbf{v} = \begin{bmatrix} 6 & -1 & 10 \end{bmatrix}^T$

    (c) $\mathbf{v} = \begin{bmatrix} 16 & 8 & -17 \end{bmatrix}^T$

13. Write down the change-of-basis matrix $[\mathcal{B}]_{\mathcal{E}}$ for the basis in question 8. Multiply $\mathbf{v}$ by $[\mathcal{B}]_{\mathcal{E}}^{-1}$ and compare your answer to that from question 8. [$\mathbb{S}$]-318

14. Write down the change-of-basis matrix $[\mathcal{B}]_{\mathcal{E}}$ for the basis in question **??**. Multiply $\mathbf{v}$ by $[\mathcal{B}]_{\mathcal{E}}^{-1}$ and compare your answer to that from question **??**.

15. Write down the change-of-basis matrix $[\mathcal{B}]_{\mathcal{E}}$ for the basis in question 9. Multiply $\mathbf{v}$ by $[\mathcal{B}]_{\mathcal{E}}^{-1}$ and compare your answer to that from question 9. [$\mathbb{A}$]-356

16. Write down the change-of-basis matrix $[\mathcal{B}]_{\mathcal{E}}$ for the basis in question 10. Multiply $\mathbf{v}$ by $[\mathcal{B}]_{\mathcal{E}}^{-1}$ and compare your answer to that from question 10.

17. Write down the change-of-basis matrix $[\mathcal{B}]_{\mathcal{E}}$ for the basis in question 11. Multiply $\mathbf{v}$ by $[\mathcal{B}]_{\mathcal{E}}^{-1}$ and compare your answer to that from question 11. [$\mathbb{A}$]-356

18. Write down the change-of-basis matrix $[\mathcal{B}]_{\mathcal{E}}$ for the basis in question 12. Multiply $\mathbf{v}$ by $[\mathcal{B}]_{\mathcal{E}}^{-1}$ and compare your answer to that from question 12.

19. Given bases $\mathcal{B} = \left\{ \begin{bmatrix} -8 \\ 7 \end{bmatrix}, \begin{bmatrix} 5 \\ -6 \end{bmatrix} \right\}$ and $\mathcal{C} = \left\{ \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix} \right\}$ of $\mathbb{R}^2$, find the change-of-basis matrix $[\mathcal{B}]_{\mathcal{C}}$. [$\mathbb{S}$]-319

20. Given bases $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \end{bmatrix} \right\}$ and $\mathcal{C} = \left\{ \begin{bmatrix} 6 \\ 1 \end{bmatrix}, \begin{bmatrix} -5 \\ 7 \end{bmatrix} \right\}$ of $\mathbb{R}^2$, find the change-of-basis matrix $[\mathcal{B}]_{\mathcal{C}}$.

21. Given bases $\mathcal{B} = \left\{ \begin{bmatrix} -7 \\ 6 \end{bmatrix}, \begin{bmatrix} 6 \\ 3 \end{bmatrix} \right\}$ and $\mathcal{C} = \left\{ \begin{bmatrix} 6 \\ -2 \end{bmatrix}, \begin{bmatrix} -3 \\ 6 \end{bmatrix} \right\}$ of $\mathbb{R}^2$, find the change-of-basis matrix $[\mathcal{B}]_{\mathcal{C}}$. [$\mathbb{A}$]-356

22. Given bases $\mathcal{B} = \left\{ \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \begin{bmatrix} 5 \\ 9 \end{bmatrix} \right\}$ and $\mathcal{C} = \left\{ \begin{bmatrix} -9 \\ 1 \end{bmatrix}, \begin{bmatrix} 6 \\ 5 \end{bmatrix} \right\}$ of $\mathbb{R}^2$, find the change-of-basis matrix $[\mathcal{B}]_{\mathcal{C}}$.

23. SageMathCell 56 Given bases $\mathcal{B} = \left\{ \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -7 \\ 7 \\ 4 \end{bmatrix}, \begin{bmatrix} -8 \\ 3 \\ 5 \end{bmatrix} \right\}$ and $\mathcal{C} = \left\{ \begin{bmatrix} 0 \\ 2 \\ 8 \end{bmatrix}, \begin{bmatrix} 8 \\ 6 \\ 8 \end{bmatrix}, \begin{bmatrix} 2 \\ -5 \\ -7 \end{bmatrix} \right\}$ of $\mathbb{R}^3$, find the change-of-basis matrix $[\mathcal{B}]_{\mathcal{C}}$. [$\mathbb{A}$]-356

24. SageMathCell 57 Given bases $\mathcal{B} = \left\{ \begin{bmatrix} 5 \\ 5 \\ -4 \end{bmatrix}, \begin{bmatrix} 8 \\ 2 \\ 7 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ -5 \end{bmatrix} \right\}$ and $\mathcal{C} = \left\{ \begin{bmatrix} 6 \\ 7 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \\ 2 \end{bmatrix}, \begin{bmatrix} 7 \\ 9 \\ 0 \end{bmatrix} \right\}$ of $\mathbb{R}^3$, find the change-of-basis matrix $[\mathcal{B}]_{\mathcal{C}}$.

25. Let $\mathbf{v} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$ and $\mathcal{B}$ and $\mathcal{C}$ be as in question 19. [$\mathbb{S}$]-319

    (a) Find $[\mathbf{v}]_{\mathcal{B}}$.

    (b) Find $[\mathbf{v}]_{\mathcal{C}}$.

    (c) Using your answer from question 19, calculate $[\mathcal{B}]_{\mathcal{C}}\,[\mathbf{v}]_{\mathcal{B}}$ and verify that it equals $[\mathbf{v}]_{\mathcal{C}}$.

26. Let $\mathbf{v} = \begin{bmatrix} 3 \\ -9 \end{bmatrix}$ and $\mathcal{B}$ and $\mathcal{C}$ be as in question 20.

    (a) Find $[\mathbf{v}]_{\mathcal{B}}$.

    (b) Find $[\mathbf{v}]_{\mathcal{C}}$.

    (c) Using your answer from question 20, calculate $[\mathcal{B}]_{\mathcal{C}}\,[\mathbf{v}]_{\mathcal{B}}$ and verify that it equals $[\mathbf{v}]_{\mathcal{C}}$.

27. Let $\mathbf{v} = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$ and $\mathcal{B}$ and $\mathcal{C}$ be as in question 21. [$\mathbb{A}$]-356

    (a) Find $[\mathbf{v}]_{\mathcal{B}}$.

    (b) Find $[\mathbf{v}]_{\mathcal{C}}$.

    (c) Using your answer from question 21, calculate $[\mathcal{B}]_{\mathcal{C}}\,[\mathbf{v}]_{\mathcal{B}}$ and verify that it equals $[\mathbf{v}]_{\mathcal{C}}$.

28. Let $\mathbf{v} = \begin{bmatrix} 8 \\ -1 \end{bmatrix}$ and $\mathcal{B}$ and $\mathcal{C}$ be as in question 22.

    (a) Find $[\mathbf{v}]_{\mathcal{B}}$.

    (b) Find $[\mathbf{v}]_{\mathcal{C}}$.

    (c) Using your answer from question 22, calculate $[\mathcal{B}]_{\mathcal{C}}\,[\mathbf{v}]_{\mathcal{B}}$ and verify that it equals $[\mathbf{v}]_{\mathcal{C}}$.

29. SageMathCell 58 Let $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ and $\mathcal{B}$ and $\mathcal{C}$ be as in question 23. [$\mathbb{A}$]-356

    (a) Find $[\mathbf{v}]_{\mathcal{B}}$.

    (b) Find $[\mathbf{v}]_{\mathcal{C}}$.

    (c) Use your answer from question 23 to calculate $[\mathcal{B}]_{\mathcal{C}}\,[\mathbf{v}]_{\mathcal{B}}$ and verify that it equals $[\mathbf{v}]_{\mathcal{C}}$.

30. 59 Let $\mathbf{v} = \begin{bmatrix} 7 \\ -6 \\ 3 \end{bmatrix}$ and $\mathcal{B}$ and C be as in question 24.
   (a) Find $[\mathbf{v}]_{\mathcal{B}}$.
   (b) Find $[\mathbf{v}]_C$.
   (c) Use your answer from question 24 to calculate $[\mathcal{B}]_C\,[\mathbf{v}]_{\mathcal{B}}$ and verify that it equals $[\mathbf{v}]_C$.

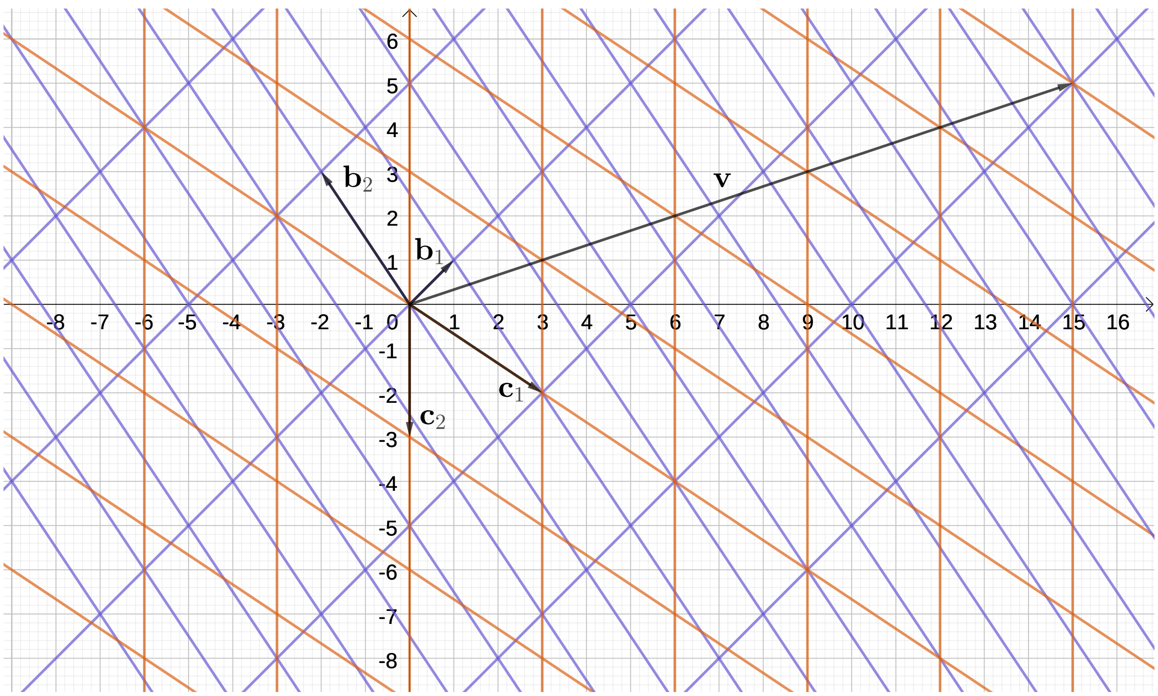
31. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ and $C = \{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$ be bases for a vector space V and suppose $\mathbf{v} = 2\mathbf{b}_1 + 5\mathbf{b}_2 - 6\mathbf{b}_3$, $\mathbf{b}_1 = 3\mathbf{c}_1 - 8\mathbf{c}_2 + 5\mathbf{c}_3$, $\mathbf{b}_2 = 6\mathbf{c}_1 - 2\mathbf{c}_2 + 9\mathbf{c}_3$, and $\mathbf{b}_3 = 7\mathbf{c}_1 + 3\mathbf{c}_2 - \mathbf{c}_3$.
   (a) Find $[\mathbf{v}]_{\mathcal{B}}$.
   (b) Find the change-of-basis matrix $[\mathcal{B}]_C$.
   (c) Find $[\mathbf{v}]_C$.

32. Given basis $\mathcal{B} = \left\{ \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \begin{bmatrix} 5 \\ 9 \end{bmatrix} \right\}$ of $\mathbb{R}^2$ and change-of-basis matrix $[\mathcal{B}]_C = \begin{bmatrix} 3 & 2 \\ -3 & 7 \end{bmatrix}$, find the basis C.

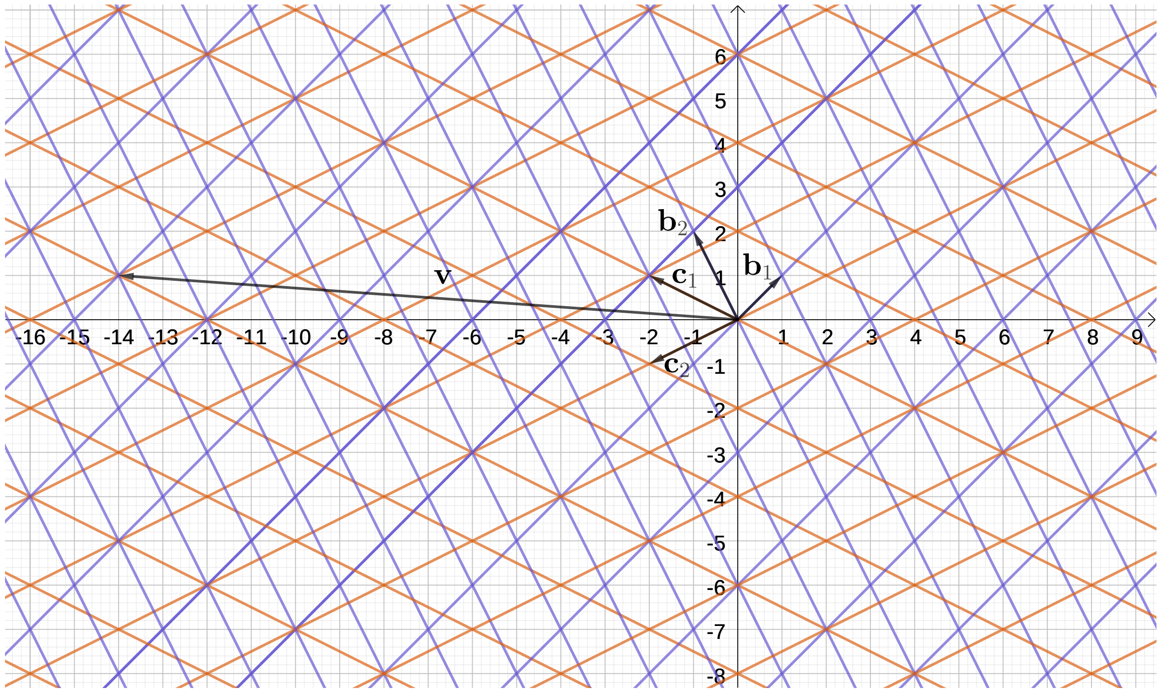
33. Given basis $\mathcal{B} = \{5 + 9t, 2 - 5t\}$ of $\mathbb{P}_1\{\mathbb{R}\}$ and change-of-basis matrix $[\mathcal{B}]_C = \begin{bmatrix} 1 & 8 \\ -7 & 1 \end{bmatrix}$, find the basis C. [$\mathbb{S}$]-319

The last few exercises connect the algebra with the geometry of coordinate systems in $\mathbb{R}^2$. The geometry and algebra of coordinates in $\mathbb{R}^n$ are connected similarly.

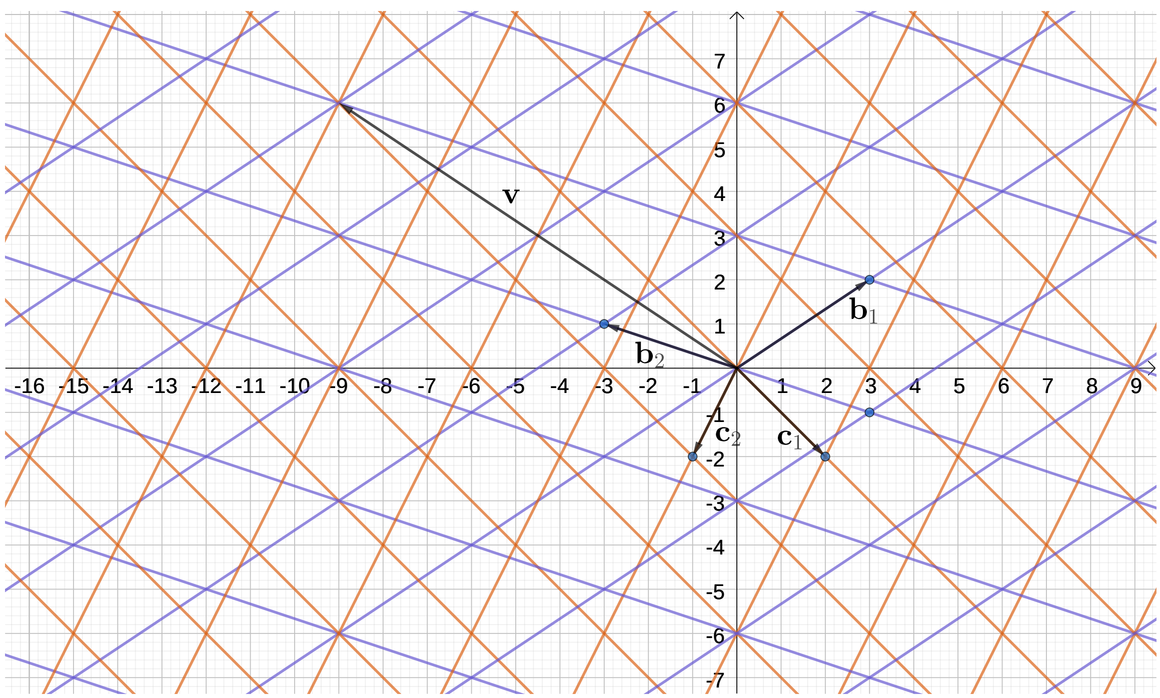
34. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $C = \{\mathbf{c}_1, \mathbf{c}_2\}$. Find $[\mathbf{v}]_{\mathcal{B}}$ and $[\mathbf{v}]_C$. [$\mathbb{S}$]-320



35. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $C = \{\mathbf{c}_1, \mathbf{c}_2\}$. Find $[\mathbf{v}]_{\mathcal{B}}$ and $[\mathbf{v}]_C$.



36. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $C = \{\mathbf{c}_1, \mathbf{c}_2\}$. Find $[\mathbf{v}]_{\mathcal{B}}$ and $[\mathbf{v}]_C$.



37. Find the change-of-basis matrix $[\mathcal{B}]_C$ and verify that $[\mathcal{B}]_C\,[\mathbf{v}]_{\mathcal{B}} = [\mathbf{v}]_C$ in question 34.

38. Find the change-of-basis matrix $[C]_{\mathcal{B}}$ and verify that $[C]_{\mathcal{B}}\,[\mathbf{v}]_C = [\mathbf{v}]_{\mathcal{B}}$ in question 35.

39. Find the change-of-basis matrix $[\mathcal{B}]_C$ and verify that $[\mathcal{B}]_C\,[\mathbf{v}]_{\mathcal{B}} = [\mathbf{v}]_C$ in question 36.

Answers

equivalent coordinate vectors The coordinate vector $\begin{bmatrix} 2 \\ 7 \\ -3 \end{bmatrix}_{\mathcal{C}}$ *means* 2 times the first vector of $\mathcal{C}$ plus 7 times the second vector of $\mathcal{C}$ minus 3 times the third vector of $\mathcal{C}$: $2\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 7\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - 3\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Similarly, $\begin{bmatrix} 6 \\ 4 \\ -3 \end{bmatrix}$ means 6 times the first vector of $\mathcal{E}$ plus 4 times the second vector of $\mathcal{E}$ minus 3 times the third vector of $\mathcal{E}$: $6I_{:,1}+4I_{:,2}-3I_{:,3}$. Finally, $\begin{bmatrix} 4 \\ 6 \\ -3 \end{bmatrix}_{\mathcal{B}}$ means 4 times the first vector of $\mathcal{B}$ plus 6 times the second vector of $\mathcal{B}$ minus 3 times the third vector of $\mathcal{B}$. Hence

$$\begin{bmatrix} 2 \\ 7 \\ -3 \end{bmatrix}_{\mathcal{C}} = 2\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 7\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - 3\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 7 \\ 7 \\ 0 \end{bmatrix} - \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 3 \end{bmatrix}$$

and

$$\begin{bmatrix} 4 \\ 6 \\ -3 \end{bmatrix}_{\mathcal{B}} = 4I_{:,2} + 6I_{:,1} - 3I_{:,3} = \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix} + \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} - 3\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ -3 \end{bmatrix}$$

5.3 Orthogonalization [4.6, 5.2]

As it is in linear algebra, determining the linear combination of basis vectors that sums to a given vector is big business in engineering and the sciences. This problem is generally aided by careful choice of basis vectors (see Legendre Polynomials, Fourier Series, and Finite Element Methods, for example). But what makes one set of basis vectors more amenable than another? To get some idea, suppose we have an arbitrary basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ of an inner product space V and an arbitrary vector $\mathbf{v}$ in V. The problem stated mathematically is to find coefficients $a_1, a_2, \dots, a_n$ such that

$$\mathbf{v} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n. \tag{5.3.1}$$

If $V = \mathbb{R}^n$, this is the vector form of a linear system. It can be solved by row reduction. If $V \neq \mathbb{R}^n$, it is less obvious how to solve. Using coordinate vectors (see section 5.2) is one way to turn (5.3.1) into a linear system, but doing so does not shed light on simplifying the process. Since V is an inner product space, though, perhaps the inner product can be leveraged instead.

Taking the inner product of both sides of (5.3.1) with $\mathbf{b}_1$ and then $\mathbf{b}_2$, and so on through $\mathbf{b}_n$ produces the following linear system.

$$\begin{aligned}
\langle \mathbf{v}, \mathbf{b}_1 \rangle &= \langle a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n, \mathbf{b}_1 \rangle \\
\langle \mathbf{v}, \mathbf{b}_2 \rangle &= \langle a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n, \mathbf{b}_2 \rangle \\
&\vdots \\
\langle \mathbf{v}, \mathbf{b}_n \rangle &= \langle a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n, \mathbf{b}_n \rangle
\end{aligned} \tag{5.3.2}$$

By inner product space properties 4 and 5, we have

$$\begin{aligned}
\langle \mathbf{v}, \mathbf{b}_1 \rangle &= a_1\langle \mathbf{b}_1, \mathbf{b}_1 \rangle + a_2\langle \mathbf{b}_2, \mathbf{b}_1 \rangle + \cdots + a_n\langle \mathbf{b}_n, \mathbf{b}_1 \rangle \\
\langle \mathbf{v}, \mathbf{b}_2 \rangle &= a_1\langle \mathbf{b}_1, \mathbf{b}_2 \rangle + a_2\langle \mathbf{b}_2, \mathbf{b}_2 \rangle + \cdots + a_n\langle \mathbf{b}_n, \mathbf{b}_2 \rangle \\
&\vdots \\
\langle \mathbf{v}, \mathbf{b}_n \rangle &= a_1\langle \mathbf{b}_1, \mathbf{b}_n \rangle + a_2\langle \mathbf{b}_2, \mathbf{b}_n \rangle + \cdots + a_n\langle \mathbf{b}_n, \mathbf{b}_n \rangle
\end{aligned}$$

which in matrix form is

$$\begin{bmatrix} \langle \mathbf{b}_1, \mathbf{b}_1 \rangle & \langle \mathbf{b}_2, \mathbf{b}_1 \rangle & \cdots & \langle \mathbf{b}_n, \mathbf{b}_1 \rangle \\ \langle \mathbf{b}_1, \mathbf{b}_2 \rangle & \langle \mathbf{b}_2, \mathbf{b}_2 \rangle & \cdots & \langle \mathbf{b}_n, \mathbf{b}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{b}_1, \mathbf{b}_n \rangle & \langle \mathbf{b}_2, \mathbf{b}_n \rangle & \cdots & \langle \mathbf{b}_n, \mathbf{b}_n \rangle \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \langle \mathbf{v}, \mathbf{b}_1 \rangle \\ \langle \mathbf{v}, \mathbf{b}_2 \rangle \\ \vdots \\ \langle \mathbf{v}, \mathbf{b}_n \rangle \end{bmatrix}. \tag{5.3.3}$$

No matter the vector, no matter the basis, and no matter the inner product space—the problem of writing a vector with respect to a given basis is reduced to solving a linear system of equations, a problem that we have studied extensively!

If that were all there were to it, it would be enough (though no better than using coordinate vectors). Note that (5.3.1) is a linear system of n equations in n unknowns, and so is (5.3.3). Is one really better than the other? This section began with the promise that careful choice of basis would help. Since the process of row reduction involves producing zeros above and below the pivots, starting with some zeros in these entries of the coefficient matrix of (5.3.3) would be beneficial. For example, if $\langle \mathbf{b}_1, \mathbf{b}_2 \rangle$ were zero, it would put zeros in the 1,2-entry and the 2,1-entry. More generally, if $\langle \mathbf{b}_i, \mathbf{b}_j \rangle$ were zero, it would put zeros in the i,j-entry and the j,i-entry. The more orthogonal pairs of basis vectors (zero inner products between basis vectors) the better. Thinking greedily, if $\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0$ for all pairs i,j, $i \neq j$, the system reads

$$\begin{bmatrix} \langle \mathbf{b}_1, \mathbf{b}_1 \rangle & 0 & \cdots & 0 \\ 0 & \langle \mathbf{b}_2, \mathbf{b}_2 \rangle & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle \mathbf{b}_n, \mathbf{b}_n \rangle \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \langle \mathbf{v}, \mathbf{b}_1 \rangle \\ \langle \mathbf{v}, \mathbf{b}_2 \rangle \\ \vdots \\ \langle \mathbf{v}, \mathbf{b}_n \rangle \end{bmatrix} \tag{5.3.4}$$

and has solution $a_1 = \frac{\langle \mathbf{v}, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle}$, $a_2 = \frac{\langle \mathbf{v}, \mathbf{b}_2 \rangle}{\langle \mathbf{b}_2, \mathbf{b}_2 \rangle}$,. . .,$a_n = \frac{\langle \mathbf{v}, \mathbf{b}_n \rangle}{\langle \mathbf{b}_n, \mathbf{b}_n \rangle}$. Better yet, if $\langle \mathbf{b}_1, \mathbf{b}_1 \rangle = \langle \mathbf{b}_2, \mathbf{b}_2 \rangle = \cdots = \langle \mathbf{b}_n, \mathbf{b}_n \rangle = 1$, the solution is simply $a_1 = \langle \mathbf{v}, \mathbf{b}_1 \rangle$, $a_2 = \langle \mathbf{v}, \mathbf{b}_2 \rangle$,. . .,$a_n = \langle \mathbf{v}, \mathbf{b}_n \rangle$—the coefficients are just the inner products of $\mathbf{v}$ with each of the basis vectors.

The question then turns to (a) establishing bases within which pairs of vectors are orthogonal, and (b) possibly ensuring all their norms (inner products $\langle \mathbf{b}_i, \mathbf{b}_i \rangle$) are one. The prototypical example of such a basis is the standard basis $\mathcal{E} = \{I_{:,1}, I_{:,2}, \ldots, I_{:,n}\}$ with inner product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v}$ (the dot product, after which inner products were modeled). Can you verify that the inner product of every pair of distinct vectors in $\mathcal{E}$ is zero and that the norm of each vector in $\mathcal{E}$ is one? Answer on page 182. The standard basis $\mathcal{E} = \{1, t, t^2\}$ of $\mathbb{P}_2(\mathbb{R})$ (see page 127) with inner product (4.6.2) does not have these properties. Can you identify at least one violation? Answer on page 182.

The SageMath output below demonstrates a process for taking any basis of $\mathbb{R}^3$ and modifying it so that all pairs of vectors are orthogonal, a process called orthogonalization.

```
Basis B:
------------
(5, -5, 0)
(7, -1, 9)
(-8, 3, 3)

Orthogonal Basis C:
-------------------
w1 = (5, -5, 0)
w2 = (7, -1, 9) - (4, -4, 0) = (3, 3, 9)
w3 = (-8, 3, 3) - (-11/2, 11/2, 0) - (4/11, 4/11, 12/11) = (-63/22, -63/22, 21/11)
```

Follow this SageCell link to generate random examples. Can you verify that all pairs of distinct vectors among

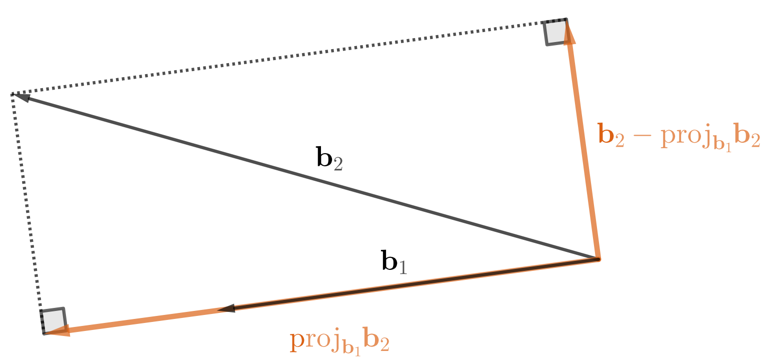
$$\begin{bmatrix} 5 \\ -5 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 9 \end{bmatrix}, \begin{bmatrix} -\frac{63}{22} \\ -\frac{63}{22} \\ \frac{21}{11} \end{bmatrix}$$

from the SageMath snapshot are orthogonal? Answer on page 181. The first vector of the original basis is taken as the first vector of the orthogonal basis. The second vector of the original basis minus a particular vector is taken as the second vector of the orthogonal basis. The third vector of the original basis minus two particular vectors is taken as the third vector of the orthogonal basis. But what vectors ought to be subtracted? A clever observation will answer the question.

Given any nonzero vectors $\mathbf{b}_1$ and $\mathbf{b}_2$,

$$\begin{aligned}
\left\langle \mathbf{b}_1, \mathbf{b}_2 - \frac{\langle \mathbf{b}_2, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle}\mathbf{b}_1 \right\rangle &= \langle \mathbf{b}_1, \mathbf{b}_2 \rangle - \left\langle \mathbf{b}_1, \frac{\langle \mathbf{b}_2, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle}\mathbf{b}_1 \right\rangle \\
&= \langle \mathbf{b}_1, \mathbf{b}_2 \rangle - \frac{\langle \mathbf{b}_2, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle} \langle \mathbf{b}_1, \mathbf{b}_1 \rangle \\
&= \langle \mathbf{b}_1, \mathbf{b}_2 \rangle - \langle \mathbf{b}_2, \mathbf{b}_1 \rangle \\
&= 0 \qquad (5.3.5)
\end{aligned}$$

so $\mathbf{b}_1$ and $\mathbf{b}_2 - \frac{\langle \mathbf{b}_2, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle}\mathbf{b}_1$ are always orthogonal (even if $\mathbf{b}_1$ and $\mathbf{b}_2$ are not). This applies in any inner product space, not just $\mathbb{R}^n$. The term $\frac{\langle \mathbf{b}_2, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle}\mathbf{b}_1$ is called the component of $\mathbf{b}_2$ in the $\mathbf{b}_1$ direction or the **orthogonal projection** of $\mathbf{b}_2$ onto $\mathbf{b}_1$, and is often denoted $\text{proj}_{\mathbf{b}_1}\mathbf{b}_2$. Subtracting this term from $\mathbf{b}_2$ removes the component of $\mathbf{b}_2$ in the $\mathbf{b}_1$ direction, leaving only the component of $\mathbf{b}_2$ orthogonal to $\mathbf{b}_1$ (not in the direction of $\mathbf{b}_1$). In $\mathbb{R}^2$, this is seen geometrically in the following diagram.



A set of three nonzero vectors, $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$, can be orthogonalized by extending the process to $\mathbf{b}_3$. Its components in the directions of both $\mathbf{b}_1$ and $\mathbf{b}_2 - \frac{\langle \mathbf{b}_2, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle}\mathbf{b}_1$ will need to be subtracted. Letting $\mathbf{w}_1 = \mathbf{b}_1$ and $\mathbf{w}_2 = \mathbf{b}_2 - \operatorname{proj}_{\mathbf{w}_1}\mathbf{b}_2$, $\mathbf{w}_3 = \mathbf{b}_3 - \operatorname{proj}_{\mathbf{w}_1}\mathbf{b}_3 - \operatorname{proj}_{\mathbf{w}_2}\mathbf{b}_3$. For larger sets, the process continues recursively.

$$\begin{aligned} \mathbf{w}_1 &= \mathbf{b}_1 \\ \mathbf{w}_j &= \mathbf{b}_j - \operatorname{proj}_{\mathbf{w}_1}\mathbf{b}_j - \operatorname{proj}_{\mathbf{w}_2}\mathbf{b}_j - \cdots - \operatorname{proj}_{\mathbf{w}_{j-1}}\mathbf{b}_j, \quad j = 2, 3, \ldots, n. \end{aligned} \tag{5.3.6}$$

The process as described by (5.3.6) is called **orthogonalization**, or Gram-Schmidt orthogonalization.

A few details have thus far been overlooked. For one, formula (5.3.6) only works if the denominators $\langle \mathbf{w}_1, \mathbf{w}_1 \rangle$, $\langle \mathbf{w}_2, \mathbf{w}_2 \rangle$, …, $\langle \mathbf{w}_{n-1}, \mathbf{w}_{n-1} \rangle$ of the projections are all nonzero. That is, the vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{n-1}$ are nonzero. Can you provide an argument that $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ being linearly independent assures $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{n-1}$ are nonzero? HINT: Show that if $\mathbf{w}_j = \mathbf{0}$ for some j, then $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ is linearly dependent. Answer on page 182. For two, the discussion at the beginning of the section establishes that (5.3.1) implies (5.3.3), but what we have been relying on is the converse, that (5.3.3) implies (5.3.1). To start resolving this issue, can you show that if $\langle \mathbf{w}, \mathbf{b}_1 \rangle = \langle \mathbf{w}, \mathbf{b}_2 \rangle = \cdots = \langle \mathbf{w}, \mathbf{b}_n \rangle = 0$ for some vector $\mathbf{w} \in V$, then $\mathbf{w} = \mathbf{0}$? Answer on page 182. This means the zero vector is orthogonal to (has inner product zero with) every vector in a basis and in fact is the only such vector. This fact plays a prominent role in this discussion. For three, the span of the orthogonalized vectors is the same as the span of the original vectors. This is a particularly important feature of the process if you are orthogonalizing the basis of a subspace. See crumpet 24 for resolutions of these last two issues.

Crumpet 24: Details of the Process of Orthogonalization

To show that that (5.3.3) implies (5.3.1), we follow the steps establishing that (5.3.1) implies (5.3.3) backward. The same properties of inner product spaces that gave us that (5.3.2) implies (5.3.3) work in reverse, giving us (5.3.3) implies (5.3.2). However, the implication from (5.3.2) to (5.3.1) is not as straightforward. Assuming (5.3.2) we need to show that $\mathbf{v} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n$. Moving everything to the lefthand side in (5.3.2),

$$\begin{aligned} \langle \mathbf{v}, \mathbf{b}_1 \rangle - \langle a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n, \mathbf{b}_1 \rangle &= 0 \\ \langle \mathbf{v}, \mathbf{b}_2 \rangle - \langle a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n, \mathbf{b}_2 \rangle &= 0 \\ &\vdots \\ \langle \mathbf{v}, \mathbf{b}_n \rangle - \langle a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n, \mathbf{b}_n \rangle &= 0 \end{aligned}$$

which implies

$$\begin{aligned} \langle \mathbf{v} - (a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n), \mathbf{b}_1 \rangle &= 0 \\ \langle \mathbf{v} - (a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n), \mathbf{b}_2 \rangle &= 0 \\ &\vdots \\ \langle \mathbf{v} - (a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n), \mathbf{b}_n \rangle &= 0 \end{aligned}$$

so $\mathbf{v} - (a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n)$ is orthogonal to each basis vector. As shown in "zero inner products" on page 182, this means $\mathbf{v} - (a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n) = \mathbf{0}$ and therefore $\mathbf{v} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n$. This settles issue two.

To show that the span of a set of vectors is not changed by orthogonalization, we show something stronger: that $\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_j\} = \operatorname{span}\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_j\}$ for all $j = 1, 2, \ldots, n$. Let $W_j = \operatorname{span}\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_j\}$, making W_j a vector space with dimension j. Each $\mathbf{w}_j$ is by construction a linear combination of $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_j\}$ and therefore in W_j. Of course $W_1 = \operatorname{span}\{\mathbf{b}_1\} \subset W_2 = \operatorname{span}\{\mathbf{b}_1, \mathbf{b}_2\} \subset \cdots \subset W_j = \operatorname{span}\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_j\}$ so $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_j$ are all in W_j. If we knew that $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_j$ were linearly independent, we would have j linearly independent vectors in a j-dimensional vector space, making $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_j\}$ a basis for W_j. Being a basis, $\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_j\}$ would equal W_j and we would be done.

Direct computation shows that a set of nonzero vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ in which every pair of distinct vectors is orthogonal is linearly independent. Suppose $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$ for scalars $c_1, c_2, \dots, c_p$. Then

$$\begin{aligned} 0 = \langle \mathbf{0}, \mathbf{v}_1 \rangle &= \left\langle (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p), \mathbf{v}_1 \right\rangle \\ &= \langle c_1\mathbf{v}_1, \mathbf{v}_1 \rangle + \langle c_2\mathbf{v}_2, \mathbf{v}_1 \rangle + \cdots + \left\langle c_p\mathbf{v}_p, \mathbf{v}_1 \right\rangle \\ &= c_1 \langle \mathbf{v}_1, \mathbf{v}_1 \rangle + c_2 \langle \mathbf{v}_2, \mathbf{v}_1 \rangle + \cdots + c_p \left\langle \mathbf{v}_p, \mathbf{v}_1 \right\rangle \\ &= c_1 \langle \mathbf{v}_1, \mathbf{v}_1 \rangle \end{aligned}$$

(see exercise 19a in section 4.6) since $\left\langle \mathbf{v}_i, \mathbf{v}_j \right\rangle = 0$ whenever $i \neq j$ (every pair of distinct vectors is orthogonal). But $\mathbf{v}_1$ is a nonzero vector, so $\langle \mathbf{v}_1, \mathbf{v}_1 \rangle \neq 0$ (inner product property 2). Hence c_1 must be zero. Similarly $c_2, c_3, \dots, c_p$ must also be zero.

It remains to show that every pair of distinct vectors in $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_j\}$ is orthogonal. Every pair of distinct vectors in $\{\mathbf{w}_1\}$ is orthogonal (vacuously since there are no distinct pairs in the set). By (5.3.5), $\mathbf{w}_1$ and $\mathbf{w}_2$ are orthogonal so every pair of distinct vectors in $\{\mathbf{w}_1, \mathbf{w}_2\}$ is orthogonal. Now suppose every pair of distinct vectors in $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{j-1}\}$ is orthogonal for some $j \geq 3$. If $1 \leq i < j$, then

$$\begin{aligned} \left\langle \mathbf{w}_i, \mathbf{w}_j \right\rangle &= \left\langle \mathbf{w}_i, \mathbf{b}_j - \operatorname{proj}_{\mathbf{w}_1}\mathbf{b}_j - \operatorname{proj}_{\mathbf{w}_2}\mathbf{b}_j - \cdots - \operatorname{proj}_{\mathbf{w}_{j-1}}\mathbf{b}_j \right\rangle \\ &= \left\langle \mathbf{w}_i, \mathbf{b}_j - \frac{\langle \mathbf{b}_j, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{b}_j, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 - \cdots - \frac{\langle \mathbf{b}_j, \mathbf{w}_{j-1} \rangle}{\langle \mathbf{w}_{j-1}, \mathbf{w}_{j-1} \rangle}\mathbf{w}_{j-1} \right\rangle \\ &= \left\langle \mathbf{w}_i, \mathbf{b}_j \right\rangle - \left\langle \mathbf{w}_i, \frac{\langle \mathbf{b}_j, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 \right\rangle - \left\langle \mathbf{w}_i, \frac{\langle \mathbf{b}_j, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 \right\rangle - \cdots - \left\langle \mathbf{w}_i, \frac{\langle \mathbf{b}_j, \mathbf{w}_{j-1} \rangle}{\langle \mathbf{w}_{j-1}, \mathbf{w}_{j-1} \rangle}\mathbf{w}_{j-1} \right\rangle \\ &= \left\langle \mathbf{w}_i, \mathbf{b}_j \right\rangle - \frac{\langle \mathbf{b}_j, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle} \langle \mathbf{w}_i, \mathbf{w}_1 \rangle - \frac{\langle \mathbf{b}_j, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle} \langle \mathbf{w}_i, \mathbf{w}_2 \rangle - \cdots - \frac{\langle \mathbf{b}_j, \mathbf{w}_{j-1} \rangle}{\langle \mathbf{w}_{j-1}, \mathbf{w}_{j-1} \rangle} \left\langle \mathbf{w}_i, \mathbf{w}_{j-1} \right\rangle \\ &= \left\langle \mathbf{w}_i, \mathbf{b}_j \right\rangle - \frac{\langle \mathbf{b}_j, \mathbf{w}_i \rangle}{\langle \mathbf{w}_i, \mathbf{w}_i \rangle} \langle \mathbf{w}_i, \mathbf{w}_i \rangle \\ &= \left\langle \mathbf{w}_i, \mathbf{b}_j \right\rangle - \left\langle \mathbf{w}_i, \mathbf{b}_j \right\rangle \\ &= 0 \end{aligned}$$

since $\langle \mathbf{w}_i, \mathbf{w}_k \rangle = 0$ whenever $k \neq i$. By induction, every pair of distinct vectors in $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_j\}$ is orthogonal.

A set of vectors in which every pair of distinct vectors is orthogonal is called an **orthogonal set**. A vector with norm 1 is called a **unit vector**. If each vector in an orthogonal set is a unit vector, it is called an **orthonormal set**. If all the vectors in an orthogonal set are scaled to have norm 1 (are normalized), the orthogonal set becomes an orthonormal set with the same span.

Returning to the original question of writing vectors as linear combinations of basis elements, we see that if $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ is an orthogonal basis, then (5.3.3) reduces to

$$a_i = \frac{\langle \mathbf{v}, \mathbf{b}_i \rangle}{\langle \mathbf{b}_i, \mathbf{b}_i \rangle}, \quad i = 1, 2, \dots, n.$$

In words, writing a vector as a linear combination of orthogonal basis vectors amounts to projecting the vector onto each of the basis vectors. As a formula, if $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ is an orthogonal basis and $\mathbf{v}$ is an arbitrary vector in an inner product space, then

$$\mathbf{v} = \left(\operatorname{proj}_{\mathbf{b}_1}\mathbf{v}\right) + \left(\operatorname{proj}_{\mathbf{b}_2}\mathbf{v}\right) + \cdots + \left(\operatorname{proj}_{\mathbf{b}_n}\mathbf{v}\right).$$

In terms of coordinate vectors,

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} \frac{\langle \mathbf{v}, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle} \\ \frac{\langle \mathbf{v}, \mathbf{b}_2 \rangle}{\langle \mathbf{b}_2, \mathbf{b}_2 \rangle} \\ \vdots \\ \frac{\langle \mathbf{v}, \mathbf{b}_n \rangle}{\langle \mathbf{b}_n, \mathbf{b}_n \rangle} \end{bmatrix}$$

Key Concepts

unit vector vector with norm 1.

orthogonal set set in which every pair of distinct vectors is orthogonal.

orthonormal set orthogonal set of unit vectors.

orthogonal projection (of one vector onto another) $\text{proj}_{\mathbf{u}}\mathbf{v} = \frac{\langle \mathbf{v},\mathbf{u}\rangle}{\langle \mathbf{u},\mathbf{u}\rangle}\mathbf{u}$.

orthogonal basis basis that is an orthogonal set. If $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ is an orthogonal basis and $\mathbf{v}$ is an arbitrary vector in an inner product space, then

$$\mathbf{v} = \left(\text{proj}_{\mathbf{b}_1}\mathbf{v}\right) + \left(\text{proj}_{\mathbf{b}_2}\mathbf{v}\right) + \cdots + \left(\text{proj}_{\mathbf{b}_n}\mathbf{v}\right).$$

normalize scale a vector to meet a certain criterion. Often this means scaling so the norm is one.

(Gram-Schmidt) orthogonalization given a linearly independent set $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$, the set $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ defined by

$$\begin{aligned} \mathbf{w}_1 &= \mathbf{b}_1 \\ \mathbf{w}_j &= \mathbf{b}_j - \text{proj}_{\mathbf{w}_1}\mathbf{b}_j - \text{proj}_{\mathbf{w}_2}\mathbf{b}_j - \cdots - \text{proj}_{\mathbf{w}_{j-1}}\mathbf{b}_j, \quad j = 2, 3, \ldots, n. \end{aligned}$$

has the property that, for $j = 1, 2, \ldots, n$, $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_j\}$ is an orthogonal basis for $\text{span}\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_j\}$.

orthogonality to a basis The only vector orthogonal to every vector of a basis is $\mathbf{0}$.

orthogonal sets and linear independence an orthogonal set of nonzero vectors is linearly independent.

Exercises

For all exercises, the inner product space is the dot product on $\mathbb{R}^n$ unless specified otherwise.

1. Is $\mathbf{v}$ a unit vector? If not, normalize it.
   (a) $\mathbf{v} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \end{bmatrix}$
   (b) $\mathbf{v} = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{bmatrix}$
   (c) $\mathbf{v} = \begin{bmatrix} \frac{15}{19} & \frac{8}{19} \end{bmatrix}$ [𝕊]-321
   (d) $\mathbf{v} = \begin{bmatrix} \frac{1}{20} & -\frac{9}{80} & -\frac{5}{16} \end{bmatrix}$
   (e) $\mathbf{v} = \begin{bmatrix} -\frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix}$ [𝔸]-356
   (f) $\mathbf{v} = \begin{bmatrix} \frac{1}{6} & \frac{1}{2} & -\frac{1}{2} & \frac{2}{3} \end{bmatrix}$ [𝔸]-356
   (g) $\mathbf{v} = \begin{bmatrix} \frac{2}{9} & \frac{5}{9} & \frac{2}{3} & \frac{4}{9} \end{bmatrix}$

2. Determine whether S is orthogonal.
   (a) $S = \left\{ \begin{bmatrix} -2 \\ 9 \end{bmatrix}, \begin{bmatrix} 18 \\ 4 \end{bmatrix} \right\}$
   (b) $S = \left\{ \begin{bmatrix} 7 \\ 7 \end{bmatrix}, \begin{bmatrix} -8 \\ 4 \end{bmatrix} \right\}$ [𝕊]-321
   (c) $S = \left\{ \begin{bmatrix} 8 \\ -5 \end{bmatrix}, \begin{bmatrix} -10 \\ -16 \end{bmatrix} \right\}$
   (d) $S = \left\{ \begin{bmatrix} -7 \\ 5 \end{bmatrix}, \begin{bmatrix} -10 \\ -14 \end{bmatrix}, \begin{bmatrix} 15 \\ 21 \end{bmatrix} \right\}$
   (e) $S = \left\{ \begin{bmatrix} 6 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} -4 \\ -9 \\ 5 \end{bmatrix} \right\}$ [𝔸]-356
   (f) $S = \left\{ \begin{bmatrix} 15 \\ -5 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \\ 5 \end{bmatrix} \right\}$
   (g) $S = \left\{ \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -\frac{1}{2} \\ -5 \\ 4 \end{bmatrix}, \begin{bmatrix} 16 \\ -5 \\ -\frac{17}{4} \end{bmatrix} \right\}$ [𝕊]-321
   (h) $S = \left\{ \begin{bmatrix} -2 \\ 9 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 4 \end{bmatrix}, \begin{bmatrix} 24 \\ 13 \\ -11 \end{bmatrix} \right\}$
   (i) $S = \left\{ \begin{bmatrix} -7 \\ 0 \\ 9 \end{bmatrix}, \begin{bmatrix} 9 \\ 8 \\ 7 \end{bmatrix}, \begin{bmatrix} 4 \\ -8 \\ 4 \end{bmatrix}, \begin{bmatrix} 18 \\ 10 \\ 2 \end{bmatrix} \right\}$ [𝔸]-356

3. Repeat question 2 in $\mathbb{R}^n$ with inner product

$$\langle \mathbf{u}, \mathbf{v}\rangle = u_1v_1 + 2u_2v_2 + \cdots + nu_nv_n$$

   where $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$. [𝕊]-321 [𝔸]-356

4. Is it possible for a pair of nonzero vectors to be orthogonal in multiple inner product spaces?

5. In the inner product space $\mathbb{P}_2(\mathbb{R})$, with inner product $\langle , \rangle : \mathbb{P}_2(\mathbb{R}) \times \mathbb{P}_2(\mathbb{R}) \to \mathbb{R}$,

$$\langle p, q\rangle = p(0)q(0) + p(1)q(1) + p(2)q(2),$$

   which of the following polynomials are orthogonal to $t^2 - 2t$?

(a) $t^2 + 6t - 7$

(b) $3t^2 - t - 2$ [𝕊]-321

(c) $6t^2 - 11t + 5$

(d) $4t^2 - 8t + 3$ [𝔸]-356

(e) $2t^2 - t - 3$

6. Let $\mathbf{u} = \begin{bmatrix} 2 & 3 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 4 & -3 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} -6 & 4 \end{bmatrix}$. Find the orthogonal projection.

(a) $\text{proj}_{\mathbf{u}}\mathbf{v}$ [𝕊]-321

(b) $\text{proj}_{\mathbf{v}}\mathbf{w}$

(c) $\text{proj}_{\mathbf{u}}\mathbf{w}$ [𝔸]-356

(d) $\text{proj}_{\mathbf{w}}\mathbf{u}$

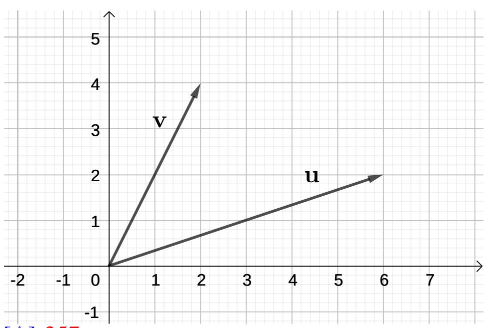
(e) $\text{proj}_{\mathbf{v}}\mathbf{u}$ [𝔸]-356

(f) $\text{proj}_{\mathbf{w}}\mathbf{v}$

7. Sketch the projection of question 6 and the two vectors involved in the projection on the same set of axes. [𝔸]-356

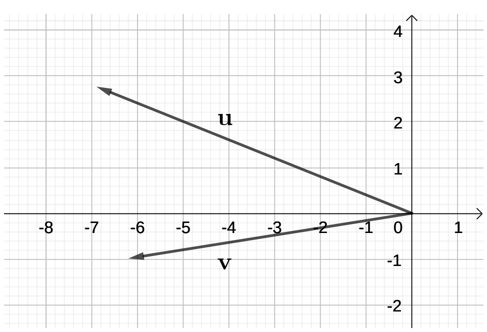
8. Approximate $\text{proj}_{\mathbf{u}}\mathbf{v}$.

(a)

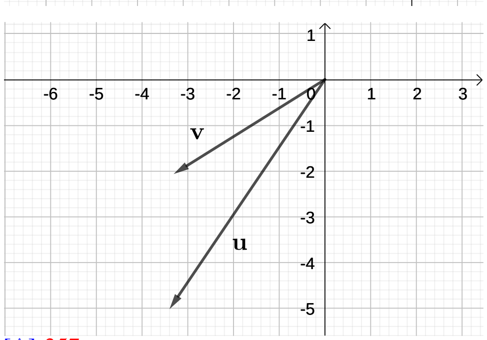


[𝔸]-357

(b)

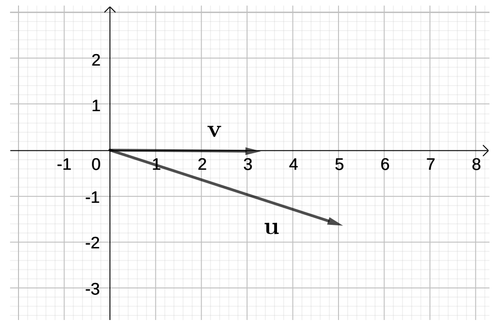


(c)



[𝔸]-357

(d)



9. Redo question 8 approximating $\text{proj}_{\mathbf{v}}\mathbf{u}$ instead. [𝔸]-357

10. True or false?

(a) In any vector space, a pair of orthogonal vectors is also perpendicular.

(b) If two vectors in $\mathbb{R}^n$ are orthogonal relative to one inner product, then they are orthogonal relative to all inner products.

(c) The zero vector is orthogonal to all vectors.

(d) Any five vectors in $\mathbb{R}^6$ can be orthogonalized (to form a set of five orthogonal vectors).

11. $\mathcal{B}$ is an orthogonal basis of $\mathbb{R}^n$. Find $[\mathbf{v}]_{\mathcal{B}}$.

(a) $\mathcal{B} = \left\{ \begin{bmatrix} -9 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}$; $\mathbf{v} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

(b) $\mathcal{B} = \left\{ \begin{bmatrix} -2 \\ -5 \end{bmatrix}, \begin{bmatrix} 5 \\ -2 \end{bmatrix} \right\}$; $\mathbf{v} = \begin{bmatrix} 8 \\ 9 \end{bmatrix}$ [𝔸]-357

(c) $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} 6 \\ 2 \end{bmatrix} \right\}$; $\mathbf{v} = \begin{bmatrix} -15 \\ 8 \end{bmatrix}$

(d) $\mathcal{B} = \left\{ \begin{bmatrix} 3 \\ 8 \\ 5 \end{bmatrix}, \begin{bmatrix} 25 \\ -15 \\ 9 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix} \right\}$; $\mathbf{v} = \begin{bmatrix} 18 \\ 5 \\ -1 \end{bmatrix}$ [𝕊]-321

(e) $\mathcal{B} = \left\{ \begin{bmatrix} -3 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \right\}$; $\mathbf{v} = \begin{bmatrix} 8 \\ 3 \\ 1 \end{bmatrix}$

12. Find an orthogonal set with the same span.

(a) $\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 \\ -7 \end{bmatrix} \right\}$ [𝕊]-321

(b) $\left\{ \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 6 \end{bmatrix} \right\}$

(c) $\left\{ \begin{bmatrix} 8 \\ -9 \end{bmatrix}, \begin{bmatrix} 9 \\ 3 \end{bmatrix} \right\}$ [𝔸]-357

(d) $\left\{ \begin{bmatrix} -7 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ -5 \end{bmatrix} \right\}$

(e) $\left\{ \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ -4 \\ 1 \end{bmatrix} \right\}$ [𝔸]-357

(f) $\left\{ \begin{bmatrix} 5 \\ -4 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} \right\}$

(g) $\left\{ \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -5 \end{bmatrix}, \begin{bmatrix} 10 \\ 2 \\ 5 \end{bmatrix} \right\}$ [𝔸]-357

(h) SageMathCell 60
$\left\{ \begin{bmatrix} -15 \\ -2 \\ 28 \end{bmatrix}, \begin{bmatrix} 13 \\ 26 \\ 8 \end{bmatrix}, \begin{bmatrix} -18 \\ 3 \\ 16 \end{bmatrix} \right\}$

(i) SageMathCell 61
$\left\{ \begin{bmatrix} -2 \\ -4 \\ 7 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 9 \\ 9 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 9 \\ -4 \end{bmatrix} \right\}$ [𝔸]-357

13. Orthonormalize (normalize the vectors resulting from the orthogonalization process as they are computed).

(a) $\left\{ \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -5 \\ 10 \end{bmatrix} \right\}$ [𝕊]-322

(b) $\left\{ \begin{bmatrix} 12 \\ 5 \end{bmatrix}, \begin{bmatrix} 7 \\ 17 \end{bmatrix} \right\}$

(c) SageMathCell 62 $\left\{ \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -5 \end{bmatrix}, \begin{bmatrix} 10 \\ 2 \\ 5 \end{bmatrix} \right\}$
[𝔸]-357

(d) SageMathCell 63 $\left\{ \begin{bmatrix} -8 \\ -9 \\ 8 \end{bmatrix}, \begin{bmatrix} 8 \\ 7 \\ 3 \end{bmatrix}, \begin{bmatrix} 6 \\ 3 \\ 5 \end{bmatrix} \right\}$

14. Is the orthogonal set from question 12 a basis for $\mathbb{R}^n$ (for any n)? [𝔸]-357

15. Scale the vectors in the orthogonalization of question 12 to find an orthonormal set. [𝔸]-357

16. Find an orthogonal set with the same span as S from question 2 and normalize each vector so it becomes an orthonormal set. [𝕊]-322 [𝔸]-358

17. Find an orthogonal set with the same span as S from question 2 relative to the inner product of question 3 and normalize each vector so it becomes an orthonormal set. [𝕊]-323 [𝔸]-358

18. The given set S is linearly dependent so we should not expect orthogonalization to lead to an orthogonal basis for span$S$. Orthogonalize anyway and explain what happens. For which j is the set $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_j\}$ (the result of orthogonalizing the first j elements of S) an orthogonal basis for the span of the first j elements of S?

(a) $S = \left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 6 \\ 9 \end{bmatrix} \right\}$

(b) $S = \left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$

(c) $S = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -5 \\ 7 \\ -3 \end{bmatrix}, \begin{bmatrix} 7 \\ -3 \\ 5 \end{bmatrix}, \begin{bmatrix} -5 \\ -6 \\ 7 \end{bmatrix} \right\}$

19. Orthogonalize $\left\{1, t, t^2\right\}$ in $\mathbb{P}_2(\mathbb{R})$ with inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\,dx.$$

Then scale each one so its graph passes through $(1,1)$. The resulting functions are the first three Legendre polynomials.

20. Orthogonalize $\left\{1, t, t^2\right\}$ in $\mathbb{P}_2(\mathbb{R})$ with inner product

$$\langle p, q \rangle = p(0)q(0) + p(1)q(1) + p(2)q(2).$$

[𝔸]-358

21. Explain why an orthogonal set is a basis for its span.

22. $\mathcal{B} = \{\sin t, \sin 2t, \sin 3t, \sin 4t\}$ is an orthogonal set in $C([-\pi, \pi])$ with inner product

$$\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)g(x)\,dx$$

and is therefore a basis for its span, W. Let $h(t) = t$ and

(a) calculate $\operatorname{proj}_{\sin kt} h$ for $k = 1, 2, 3, 4$; and

(b) graph h and $\operatorname{proj}_{\sin t} h + \operatorname{proj}_{\sin 2t} h + \operatorname{proj}_{\sin 3t} h + \operatorname{proj}_{\sin 4t} h$ on the same set of axes.

23. Repeat question 22 with

$$h(t) = \begin{cases} -1 & t < 0 \\ 1 & t \geq 0 \end{cases}.$$

Answers

orthogonal pairs The three possible dot products are all zero:

$$\begin{bmatrix} 5 \\ -5 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 3 \\ 9 \end{bmatrix} = 5(3) - 5(3) + 0(9) = 0$$

$$\begin{bmatrix} 5 \\ -5 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} -\frac{63}{22} \\ -\frac{63}{22} \\ \frac{21}{11} \end{bmatrix} = -5\left(\frac{63}{22}\right) + 5\left(\frac{63}{22}\right) + 0\left(\frac{21}{11}\right) = 0$$

$$\begin{bmatrix} 3 \\ 3 \\ 9 \end{bmatrix} \cdot \begin{bmatrix} -\frac{63}{22} \\ -\frac{63}{22} \\ \frac{21}{11} \end{bmatrix} = -3\left(\frac{63}{22}\right) - 3\left(\frac{63}{22}\right) + 9\left(\frac{21}{11}\right) = -\frac{189}{22} + \frac{189}{22} = 0$$

standard basis inner products For pairs of basis vectors, $i < j$,

$$\begin{aligned}\langle I_{:,i}, I_{:,j}\rangle &= I_{:,i} \bullet I_{:,j}\\ &= 0\cdot 0 + \cdots + 0\cdot 0 + \overbrace{1\cdot 0}^{i^{th}\text{ term}} + 0\cdot 0 + \cdots + 0\cdot 0 + \overbrace{0\cdot 1}^{j^{th}\text{ term}} + 0\cdot 0 + \cdots + 0\cdot 0\\ &= 0.\end{aligned}$$

For $i > j$, the same computation holds by symmetry (property 3 of inner product spaces). For the norms of basis vectors,

$$\begin{aligned}\|I_{:,i}\| &= \sqrt{I_{:,i} \bullet I_{:,i}}\\ &= \sqrt{0\cdot 0 + \cdots + 0\cdot 0 + \overbrace{1\cdot 1}^{i^{th}\text{ term}} + 0\cdot 0 + \cdots + 0\cdot 0}\\ &= 1.\end{aligned}$$

$\mathbb{P}_2(\mathbb{R})$ violation Violations are easy to come by. The inner products $\langle 1, t\rangle$ and $\langle t, t^2\rangle$ are both zero, but $\langle 1, t^2\rangle$ is not:

$$\begin{aligned}\langle 1, t^2\rangle &= \langle 1(1) + 0(t) + 0(t^2),\ 0(1) + 0(t) + 1(t^2)\rangle\\ &= \frac{2}{5}(0\cdot 1) + \frac{2}{3}(1\cdot 1) + \frac{2}{3}(0\cdot 0) + \frac{2}{3}(0\cdot 0) + 2(1\cdot 0)\\ &= \frac{2}{3}.\end{aligned}$$

None of the norms are one:[2]

$$\begin{aligned}\|1\| &= \sqrt{\langle 1, 1\rangle} = \sqrt{\frac{2}{5}(0\cdot 0) + \frac{2}{3}(1\cdot 0) + \frac{2}{3}(0\cdot 0) + \frac{2}{3}(0\cdot 1) + 2(1\cdot 1)} = \sqrt{2}\\ \|t\| &= \sqrt{\langle t, t\rangle} = \sqrt{\frac{2}{5}(0\cdot 0) + \frac{2}{3}(0\cdot 0) + \frac{2}{3}(1\cdot 1) + \frac{2}{3}(0\cdot 0) + 2(0\cdot 0)} = \sqrt{\frac{2}{3}}\\ \|t^2\| &= \sqrt{\langle t^2, t^2\rangle} = \sqrt{\frac{2}{5}(1, 1) + \frac{2}{3}(0\cdot 1) + \frac{2}{3}(0\cdot 0) + \frac{2}{3}(1\cdot 0) + 2(0\cdot 0)} = \sqrt{\frac{2}{5}}.\end{aligned}$$

orthogonalized vectors are nonzero Suppose $\mathbf{w}_j = \mathbf{0}$ for some j. Since each $\mathbf{w}_i$ is a linear combination of $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_i$

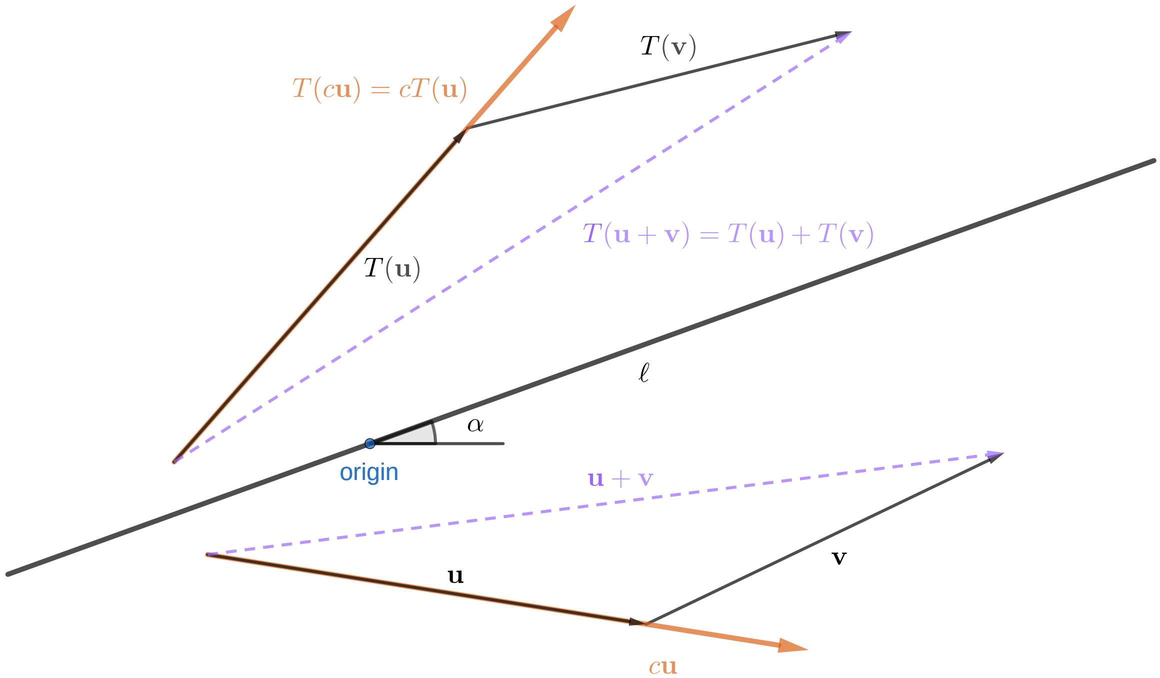
$$\begin{aligned}\mathbf{w}_j &= \mathbf{b}_j - \mathrm{proj}_{\mathbf{w}_1}\mathbf{b}_j - \mathrm{proj}_{\mathbf{w}_2}\mathbf{b}_j - \cdots - \mathrm{proj}_{\mathbf{w}_{j-1}}\mathbf{b}_j\\ &= \mathbf{b}_j + \text{some linear combination of } \mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{j-1}\end{aligned}$$

giving a nontrivial linear combination of $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_j$ that sums to $\mathbf{0}$. This implies $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_j\}$, and therefore $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$, is linearly dependent.

zero inner products Let $\mathbf{w} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n$. Then $\langle \mathbf{w}, \mathbf{b}_1\rangle = \langle \mathbf{w}, \mathbf{b}_2\rangle = \cdots = \langle \mathbf{w}, \mathbf{b}_n\rangle = 0$ means $c_1\langle \mathbf{w}, \mathbf{b}_1\rangle = c_2\langle \mathbf{w}, \mathbf{b}_2\rangle = \cdots = c_n\langle \mathbf{w}, \mathbf{b}_n\rangle = 0$. By properties 3, 4 and 5 of inner product spaces (see page 154), $\langle \mathbf{w}, c_1\mathbf{b}_1\rangle = \langle \mathbf{w}, c_2\mathbf{b}_2\rangle = \cdots = \langle \mathbf{w}, c_n\mathbf{b}_n\rangle = 0$, so $\langle \mathbf{w}, c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n\rangle = 0$. But $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n = \mathbf{w}$ so $\langle \mathbf{w}, \mathbf{w}\rangle = 0$ and by property 2, $\mathbf{w} = \mathbf{0}$.

[2] Since $1^2 = 1$, it is just as well to see that the norm squared is not 1 as in $\|1\|^2 = \langle 1, 1\rangle = \frac{2}{5}(0\cdot 0) + \frac{2}{3}(1\cdot 0) + \frac{2}{3}(0\cdot 0) + \frac{2}{3}(0\cdot 1) + 2(1\cdot 1) = 2$.

Figure 5.4.1: Reflection about the line ℓ is linear



5.4 Similarity and Diagonalization [3.7, 4.4, 5.2]

Figure 5.4.1 illustrates that reflection about an arbitrary line through the origin is a linear transformation. As we saw in section 4.4, there must therefore be a 2×2 matrix M such that $T(\mathbf{x}) = M\mathbf{x}$. In the same section, we also learned that the columns of M must satisfy $M_{:j} = T(I_{:,j})$, $j = 1, 2$, so if we knew the images of the standard basis elements, we would know M.

There is another way. Imagine rotating the plane by angle $-\alpha$ about the origin, then reflecting about the x-axis, then rotating (back) by angle α about the origin. The line ℓ would first map onto the x-axis, all vectors/points/sets in the plane would then be reflected about this image, and then the line and the reflected images would be rotated back so that ℓ would be back where it started. All vectors/points/sets and their reflections would maintain their relative positions across ℓ. In the end it would be as though the vectors/points/sets were simply reflected about line ℓ.

The composition of the two rotations and the reflection is easy enough to write down as a matrix transformation. Using the standard matrix for reflection about the x-axis (given in the discussion of section 4.4) and the standard matrix for counterclockwise rotation about the origin (from exercise 12 of section 4.4),

$$\begin{aligned} M &= \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(-\alpha) & -\sin(-\alpha) \\ \sin(-\alpha) & \cos(-\alpha) \end{bmatrix} \\ &= \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \end{aligned} \tag{5.4.1}$$

which can of course be calculated to get

$$\begin{aligned} M &= \begin{bmatrix} \cos^2\alpha - \sin^2\alpha & 2\cos\alpha\sin\alpha \\ 2\cos\alpha\sin\alpha & -(\cos^2\alpha - \sin^2\alpha) \end{bmatrix} \\ &= \begin{bmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{bmatrix}. \end{aligned} \tag{5.4.2}$$

Interestingly this form reveals that T can also be described by reflection about the x-axis followed by counterclockwise rotation by 2α about the origin. Can you see why? Answer on page 190. Standard matrices for scaling in the direction of any line and shearing along any line can be derived similarly.

Considering M in form (5.4.1) leads to a deeper perspective. As seen in equation (5.2.1), left-multiplying by an invertible matrix can always be interpreted as a change of basis. For example, left-multiplication by the matrix

$$P = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix},$$

the leftmost matrix in (5.4.1), can be interpreted as changing from the basis

$$\mathcal{B} = \left\{ \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}, \begin{bmatrix} -\sin\alpha \\ \cos\alpha \end{bmatrix} \right\}$$

to the standard basis,

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

Note that the rightmost matrix of (5.4.1) is P^{-1}. Can you verify this? Answer on page 190. Accordingly, the rightmost matrix of (5.4.1) can be interpreted as changing from the standard basis to basis $\mathcal{B}$. Altogether then, left-multiplication by M represents a change from the standard basis to basis $\mathcal{B}$, then reflection about the first basis vector in $\mathcal{B}$ (which lies along line ℓ), followed by a change of basis from $\mathcal{B}$ to the standard basis. The transformation starts with coordinates relative the the standard basis and ends with coordinates relative to the standard basis.

Taking this new perspective allows us to understand general transformations such as

$$S : \mathbb{R}^2 \to \mathbb{R}^2,\ S(\mathbf{u}) = \begin{bmatrix} -7 & -4 \\ 6 & 7 \end{bmatrix} \mathbf{u}$$

geometrically by writing the matrix as a product PAP^{-1} where the action of A is easily understood. In particular, matrices A of the form

$$\begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}$$

(diagonal matrices) are easily understood geometrically. They can be writen as the product

$$\begin{bmatrix} 1 & 0 \\ 0 & \beta \end{bmatrix} \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix},$$

which according to section 4.4 is the standard matrix of scaling by a factor of α in the direction of the first basis vector followed by scaling by a factor of β in the direction of the second basis vector. If α is negative, it incorporates a reflection about the second basis vector, and if β is negative it incorporates a reflection about the first basis vector.

Summarizing, supppose we have the standard matrix M of a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ and we want to better understand M by finding a matrix P and a diagonal matrix D such that $M = PDP^{-1}$. Right multiplying both sides by P we require $MP = PD$. The left side of this equation can be written as

$$\begin{bmatrix} MP_{:,1} & MP_{:,2} & \cdots & MP_{:,n} \end{bmatrix}$$

and the right side can be written as

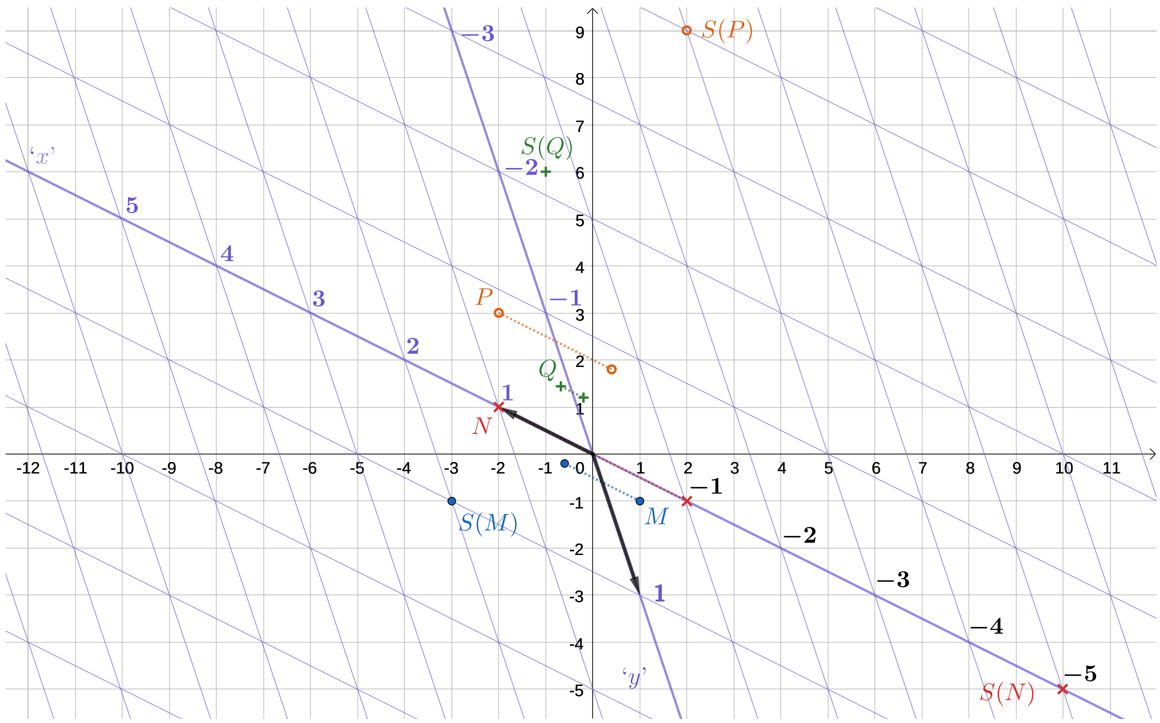
$$\begin{bmatrix} D_{1,1}P_{:,1} & D_{2,2}P_{:,2} & \cdots & D_{n,n}P_{:,n} \end{bmatrix}.$$

Equating columns of the two sides, we get

$$\begin{aligned} MP_{:,1} &= D_{1,1}P_{:,1} \\ MP_{:,2} &= D_{2,2}P_{:,2} \\ &\vdots \\ MP_{:,n} &= D_{n,n}P_{:,n}. \end{aligned}$$

The columns of P must be eigenvectors of M and the entries of D the corresponding eigenvalues!

Figure 5.4.2: Visualizing the action of S



In the particular case of $S(\mathbf{u}) = \begin{bmatrix} -7 & -4 \\ 6 & 7 \end{bmatrix}\mathbf{u}$, the eigenvalues are -5 and 5 with corresponding eigenvectors $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$ respectively. Can you provide the calculation? Answer on page 190. It follows that

$$\begin{bmatrix} -7 & -4 \\ 6 & 7 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -3 \end{bmatrix}\begin{bmatrix} -5 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} -\frac{3}{5} & -\frac{1}{5} \\ -\frac{1}{5} & -\frac{2}{5} \end{bmatrix}$$

and we see that S has the effect of reflection about the second vector of the basis $C = \left\{\begin{bmatrix} -2 \\ 1 \end{bmatrix}\begin{bmatrix} 1 \\ -3 \end{bmatrix}\right\}$ plus expansion by a factor of 5. This is because multiplication by

$$\begin{bmatrix} -5 & 0 \\ 0 & 5 \end{bmatrix}$$

has the effect of reflection about the second basis vector (rooted at the origin) plus expansion by a factor of 5. As with reflection relative to the standard basis, the reflection about the second basis vector occurs *parallel* to the first vector. Because the standard basis vectors meet at a right angle, reflection is done along lines perpendicular to the axes. If the basis vectors meet at a different angle, reflection is done along lines meeting at that same angle.

This analysis can be verified by plotting a couple of points and their images under S. For example,

$$S\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} -3 \\ -1 \end{bmatrix};\ S\left(\begin{bmatrix} -2 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 10 \\ -5 \end{bmatrix};\ S\left(\begin{bmatrix} -2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 9 \end{bmatrix};\ S\left(\begin{bmatrix} -\frac{17}{25} \\ \frac{36}{25} \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 6 \end{bmatrix} \tag{5.4.3}$$

Figure 5.4.2 shows the geometry of S with respect to both the standard basis and the basis C. The purple grid shows coordinates with respect to C. The $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ direction is the positive "x" axis (first basis vector) and the $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$ direction is the positive "y" axis (second basis vector) in these coordinates. Notice the point $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ has coordinates $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ with respect to the purple grid. In other words,

$$\begin{bmatrix} -2 \\ 1 \end{bmatrix}_{\mathcal{E}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{C}$$

They are the same point in the plane! Coordinates with respect to C make it easy to trace the action under S. $\begin{bmatrix} 1 \\ 0 \end{bmatrix}_{C}$ is first reflected about the second basis vector, making its coordinates $\begin{bmatrix} -1 \\ 0 \end{bmatrix}_{C}$ (with respect to the purple grid) and then expanded by a factor of 5, making its cordinates $\begin{bmatrix} -5 \\ 0 \end{bmatrix}_{C}$ (with respect to the purple grid). Other points can be understood similarly. The coordinates of point P are $\begin{bmatrix} -2 \\ 3 \end{bmatrix}$ with respect to the standard basis, but $\begin{bmatrix} \frac{3}{5} \\ -\frac{4}{5} \end{bmatrix}_{C}$ with respect to C (marked as P in figure 5.4.2). Can you verify this? Answer on page 190. Reflection across the second basis vector gives it coordinates $\begin{bmatrix} -\frac{3}{5} \\ -\frac{4}{5} \end{bmatrix}_{C}$ (shown in figure 5.4.2 connected to P by an orange dashed line segment crossing the "y" axis parallel to the "x" axis) and then expanding by a factor of 5 gives it coordinates $\begin{bmatrix} -3 \\ -4 \end{bmatrix}_{C}$. This point is marked as $S(P)$ in figure 5.4.2. What are the coordinates of $S(P)$ with respect to the standard basis? Answer on page 191.

Similarity

Separating the matrix calculations of the preceding discussion from their geometric interpretation, we have been examining matrices A and B related by the equation

$$B = P^{-1}AP,$$

a relation known as **similarity**. Indeed, matrices A and B are called **similar** if $B = P^{-1}AP$ (or equivalently $PBP^{-1} = A$). Consistent with the name, matrices related this way share a number of similar features.

Theorem 14. *If matrices A and B are similar, then A and B have the same (i) determinant, (ii) eigenvalues, and (iii) rank.*

(i) By theorem 8 and the fact that $\det P^{-1} = \frac{1}{\det P}$ (section 3.7), $\det B = \det(P^{-1}AP) = \det P^{-1} \cdot \det A \cdot \det P = \frac{1}{\det P} \cdot \det A \cdot \det P = \det A$.

(ii) For any value λ,

$$\begin{aligned} \det(B - \lambda I) &= \det(P^{-1}AP - \lambda I) \\ &= \det(P^{-1}AP - P^{-1}(\lambda I)P) \\ &= \det(P^{-1}(A - \lambda I)P) \\ &= \det(P^{-1}) \cdot \det(A - \lambda I) \cdot \det P \\ &= \det(A - \lambda I) \end{aligned}$$

so the characteristic equations of B and A are equal making the eigenvalues of B and A equal.

(iii) By exercises 20 and 21 of section 5.1, neither right-multiplying nor left-multiplying a matrix by an invertible matrix changes its rank, so rank of A, which equals the rank of PBP^{-1}, equals the rank of B.

Other similar features of similar matrices are explored in the exercises.

Another important feature of similar matrices is that powers of similar matrices are similar. That is, if A and B are square and similar, then A^k and B^k are similar, where the k^{th} power of M is defined by

$$M^k = \overbrace{M \cdot M \cdots M}^{k \text{ times}}$$

(analogous to powers of real numbers). To see that this is true, write $A = PBP^{-1}$ and compute

$$\begin{aligned}
A^k &= (PBP^{-1})^k \\
&= \overbrace{PBP^{-1} \cdot PBP^{-1} \cdots PBP^{-1}}^{k \text{ times}} \\
&= P\overbrace{B \cdot B \cdots B}^{k \text{ times}}P^{-1} \\
&= PB^kP^{-1}
\end{aligned}$$

This is particularly useful when A is similar to a diagonal matrix. In this case, $A = PDP^{-1}$ for a diagonal matrix D and

$$\begin{aligned}
A^k &= PD^kP^{-1} \\
&= P\begin{bmatrix} D_{1,1} & 0 & \cdots & 0 \\ 0 & D_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_{n,n} \end{bmatrix}^k P^{-1} \\
&= P\begin{bmatrix} D_{1,1}^k & 0 & \cdots & 0 \\ 0 & D_{2,2}^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_{n,n}^k \end{bmatrix} P^{-1}
\end{aligned}$$

so the difficulty in raising A to any power is commensurate with the difficulty of raising D to that power. Earlier we found that

$$\begin{bmatrix} -7 & -4 \\ 6 & 7 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -3 \end{bmatrix}\begin{bmatrix} -5 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} -\frac{3}{5} & -\frac{1}{5} \\ -\frac{1}{5} & -\frac{2}{5} \end{bmatrix}$$

so

$$\begin{aligned}
\begin{bmatrix} -7 & -4 \\ 6 & 7 \end{bmatrix}^5 &= \begin{bmatrix} -2 & 1 \\ 1 & -3 \end{bmatrix}\begin{bmatrix} -5 & 0 \\ 0 & 5 \end{bmatrix}^5\begin{bmatrix} -\frac{3}{5} & -\frac{1}{5} \\ -\frac{1}{5} & -\frac{2}{5} \end{bmatrix} \\
&= \begin{bmatrix} -2 & 1 \\ 1 & -3 \end{bmatrix}\begin{bmatrix} -3125 & 0 \\ 0 & 3125 \end{bmatrix}\begin{bmatrix} -\frac{3}{5} & -\frac{1}{5} \\ -\frac{1}{5} & -\frac{2}{5} \end{bmatrix} \\
&= \begin{bmatrix} -4375 & -2500 \\ 3750 & 4375 \end{bmatrix}.
\end{aligned}$$

While that may not be the most pleasant computation, it certainly beats computing

$$\begin{bmatrix} -7 & -4 \\ 6 & 7 \end{bmatrix}^5 = \begin{bmatrix} -7 & -4 \\ 6 & 7 \end{bmatrix}\begin{bmatrix} -7 & -4 \\ 6 & 7 \end{bmatrix}\begin{bmatrix} -7 & -4 \\ 6 & 7 \end{bmatrix}\begin{bmatrix} -7 & -4 \\ 6 & 7 \end{bmatrix}\begin{bmatrix} -7 & -4 \\ 6 & 7 \end{bmatrix}$$

directly. This property of similar matrices is at the heart of the power method for estimating eigenvalues (section 6.2), which is at the heart of Markov chain problems (section 7.2).

When a matrix M is similar to a diagonal matrix D, we say that M is **diagonalizable** and that the matrix P of $P^{-1}MP = D$ **diagonalizes** M. We saw earlier that if $P^{-1}MP = D$ then the columns of P are eigenvectors. We shall now add the observation that the eigenvectors (columns of P) must be linearly independent, a requirement for P to be invertible. On the other hand, if M is an $n \times n$ matrix admitting n linearly independent eigenvectors, then M is diagonalizable and P, a matrix whose columns are n linearly independent eigenvectors of M, diagonalizes M:

$$\begin{aligned} P^{-1}MP &= P^{-1}\begin{bmatrix} \lambda_1 P_{:,1} & \lambda_2 P_{:,2} & \cdots & \lambda_n P_{:,n} \end{bmatrix} \\ &= P^{-1}P\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}, \end{aligned}$$

which is diagonal. Altogether, an $n \times n$ matrix M is diagonalizable if and only if M has n linearly independent eigenvectors.

Key Concepts

similar matrices matrices A and B are similar if there is an invertible matrix P such that $A = P^{-1}BP$ (equivalently $PAP^{-1} = B$ or $PA = BP$).

similarity Matrices that are similar have similarity. Matrices that have similarity have the same (i) determinant, (ii) eigenvalues, and (iii) rank.

diagonalizable a matrix that is similar to a diagonal matrix.

diagonalizability an $n \times n$ matrix M is diagonalizable if and only if M has n linearly independent eigenvectors. Such M is diagonalized by any matrix whose columns are n linearly independent eigenvectors.

powers of matrices The k^{th} power of matrix M is deifined by

$$M^k = \overbrace{M \cdot M \cdots M}^{k \text{ times}}.$$

similarity and powers The k^{th} powers of similar matrices are similar.

geometry of diagonalizable matrices If M is a diagonalizable 2×2, respectively 3×3, matrix, its action on the plane, respectively space, can be understood as a scaling (and possibly reflecting) transformation relative to a basis of eigenvectors.

Exercises

1. Find a matrix P that diagonalizes M and calculate (the diagonal matrix) $P^{-1}MP$. Eigenvectors of M are given.

   (a) $M = \begin{bmatrix} 5 & 4 \\ 20 & 7 \end{bmatrix}; \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ [𝕊]-323

   (b) $M = \begin{bmatrix} 26 & 32 \\ -21 & -26 \end{bmatrix}; \begin{bmatrix} 8 \\ -7 \end{bmatrix}, \begin{bmatrix} 4 \\ -3 \end{bmatrix}$

   (c) $M = \begin{bmatrix} 1 & -2 \\ 0 & -1 \end{bmatrix}; \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ [𝔸]-358

   (d) $M = \begin{bmatrix} -17 & 6 \\ 2 & -18 \end{bmatrix}; \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \end{bmatrix}$

   (e) $M = \begin{bmatrix} -20 & 0 & 0 \\ 7 & -24 & -8 \\ -19 & 16 & 0 \end{bmatrix}; \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$ [𝔸]-358

   (f) $M = \begin{bmatrix} -12 & 100 & 30 \\ 75 & 13 & 30 \\ -75 & -100 & -117 \end{bmatrix}; \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -5 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$

   (g) $M = \begin{bmatrix} 11 & -20 & 30 \\ 8 & -17 & 30 \\ 4 & -10 & 18 \end{bmatrix}; \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ 1 \\ 2 \end{bmatrix}$ [𝕊]-323

(h) $M = \begin{bmatrix} -17 & 10 & 6 \\ -7 & 2 & 3 \\ -7 & 5 & 0 \end{bmatrix}; \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$

(i) $M = \begin{bmatrix} -39 & -28 & -52 & -20 \\ -102 & -91 & -171 & -48 \\ 68 & 56 & 107 & 32 \\ 62 & 56 & 99 & 37 \end{bmatrix};$ $\begin{bmatrix} 0 \\ 3 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ -2 \\ -2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix}$ [𝔸]-358

2. Find the eigenvalue(s) of the matrix M of question 1. [𝕊]-324 [𝔸]-358

3. Is M diagonalizable? All of its eigenvalues are given.

   (a) $M = \begin{bmatrix} 25 & 4 \\ -1 & 29 \end{bmatrix}; 27$ [𝔸]-358

   (b) $M = \begin{bmatrix} 61 & 4 \\ -25 & 41 \end{bmatrix}; 51$

   (c) $M = \begin{bmatrix} 4 & -6 \\ -9 & -11 \end{bmatrix}; 7, -14$

   (d) $M = \begin{bmatrix} 8 & 0 \\ 15 & 3 \end{bmatrix}; 8, 3$ [𝕊]-324

   (e) $M = \begin{bmatrix} 18 & -8 & 14 \\ -8 & 6 & -7 \\ -32 & 16 & -26 \end{bmatrix}; -6, 2$ [𝕊]-324

   (f) $M = \begin{bmatrix} 4 & -18 & -13 \\ 6 & 28 & 15 \\ -8 & -24 & -10 \end{bmatrix}; 10, 6$

   (g) $M = \begin{bmatrix} 15 & 14 & 18 \\ 26 & 3 & 18 \\ -26 & -14 & -29 \end{bmatrix}; 11, -11$

   (h) $M = \begin{bmatrix} 9 & 2 & 2 \\ 3 & 9 & -1 \\ -1 & -3 & 7 \end{bmatrix}; 5, 10$ [𝔸]-358

   (i) $M = \begin{bmatrix} -52 & -5 & 5 \\ 68 & -78 & 106 \\ 76 & -32 & 60 \end{bmatrix}; 28, -42, -56$

   (j) $M = \begin{bmatrix} -298 & 79 & 359 & -32 \\ 864 & -791 & -1797 & 726 \\ -800 & 631 & 1557 & -590 \\ -464 & 346 & 926 & -616 \end{bmatrix};$ $-74, 370, -296, -148$

4. Matrices A and B are similar. Find k.

   (a) $A = \begin{bmatrix} 12 & 5 \\ -11 & 10 \end{bmatrix}; B = \begin{bmatrix} 767 & k \\ -146 & -745 \end{bmatrix}$

   (b) $A = \begin{bmatrix} 5 & -2 \\ -4 & 3 \end{bmatrix}; B = \begin{bmatrix} -3 & -2 \\ 20 & k \end{bmatrix}$ [𝕊]-324

   (c) $A = \begin{bmatrix} -6 & k \\ -12 & 1 \end{bmatrix}; B = \begin{bmatrix} 3 & -10 \\ k & -8 \end{bmatrix}$

   (d) $A = \begin{bmatrix} 2 & 1 \\ k & 3 \end{bmatrix}; B = \begin{bmatrix} 3 & 5 \\ k & 2 \end{bmatrix}$ [𝔸]-358

5. Find $P \neq \mathbf{0}$ such that $PA = BP$ for the matrices of question 4. [𝕊]-324 [𝔸]-358

6. Matrices A and B are similar. Find j and k.

   (a) $A = \begin{bmatrix} j & -70 \\ k & 38 \end{bmatrix}; B = \begin{bmatrix} -3 & -4 \\ 1 & -8 \end{bmatrix}$

   (b) $A = \begin{bmatrix} j & -4 \\ -2 & k \end{bmatrix}; B = \begin{bmatrix} -9 & 2 \\ -12 & 3 \end{bmatrix}$

7. Find P such that $PAP^{-1} = B$ for the matrices of question 6.

8. Explain why A and B are not similar.

   (a) $A = \begin{bmatrix} -6 & -1 \\ 11 & -2 \end{bmatrix}; B = \begin{bmatrix} 5 & -10 \\ -4 & 8 \end{bmatrix}$ [𝕊]-325

   (b) $A = \begin{bmatrix} -2 & -8 \\ 5 & -1 \end{bmatrix}; B = \begin{bmatrix} -4 & -2 \\ 7 & 6 \end{bmatrix}$

   (c) $A = \begin{bmatrix} 37 & 9 \\ 4 & 1 \end{bmatrix}; B = \begin{bmatrix} 1 & -2 \\ -4 & 9 \end{bmatrix}$

   (d) $A = \begin{bmatrix} 3 & 4 \\ 8 & 2 \end{bmatrix}; B = \begin{bmatrix} 3 & -6 \\ 7 & -14 \end{bmatrix}$

9. What does the SageMath code do?

```
# Variables
var('a,b,c,d')
# Similar matrices
A = matrix(2,2,[9,-6,5,-8])
B = matrix(2,2,[106,-132,84,-105])
# Unknown matrix
P = matrix(2,2,[a,b,c,d])
print(A);print()
print(B);print()
# Solve for P and print solution
syseq = (A*P-P*B).coefficients()
print(solve(syseq,[a,b,c,d]))
```

10. Find all matrices P such that $A = PBP^{-1}$.

    (a) 64 $A = \begin{bmatrix} -3 & 9 \\ 9 & -2 \end{bmatrix}; B = \begin{bmatrix} -2 & 1 \\ 81 & -3 \end{bmatrix}$ [𝕊]-325

    (b) 65 $A = \begin{bmatrix} 2 & 1 \\ -7 & 2 \end{bmatrix}; B = \begin{bmatrix} 237 & 64 \\ -863 & -233 \end{bmatrix}$

    (c) 66 $A = \begin{bmatrix} 5 & 6 & 7 \\ 7 & -7 & -6 \\ 4 & 1 & -6 \end{bmatrix}; B = \begin{bmatrix} 8862 & 1601 & -33246 \\ -6634 & -1171 & 24959 \\ 2053 & 372 & -7699 \end{bmatrix}$ [𝔸]-359

    (d) 67 $A = \begin{bmatrix} -3 & 0 & 7 \\ 6 & -2 & -2 \\ 5 & 1 & -6 \end{bmatrix}; B = \begin{bmatrix} 2239 & -676 & 16323 \\ -56 & 46 & -420 \\ -315 & 94 & -2296 \end{bmatrix}$

11. What is the dimension of the solution space in question 10? [𝔸]-359

12. Describe the action of the transformation $T(\mathbf{v}) = P^{-1}MP\mathbf{v}$ as simply as you can in geometric terms.

   (a) $M = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$; $P = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}$

   (b) $M = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$; $P = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 1 \end{bmatrix}$ [𝔸]-359

   (c) $M = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$; $P = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}$ [𝔸]-359

   (d) $M = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$; $P = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$

13. Calculate M^7 using a diagonalization process. Eigenvalues of the matrix are given.

   (a) $M = \begin{bmatrix} -\frac{17}{6} & -\frac{11}{6} \\ \frac{11}{3} & \frac{8}{3} \end{bmatrix}$; $\frac{5}{6}, -1$ [𝕊]-325

   (b) $M = \begin{bmatrix} 0 & 5 \\ 1 & 4 \end{bmatrix}$; $-1, 5$

   (c) $M = \begin{bmatrix} -7 & -10 \\ 5 & 8 \end{bmatrix}$; $-2, 3$

   (d) $M = \begin{bmatrix} \frac{17}{20} & -\frac{1}{20} \\ -\frac{7}{20} & \frac{23}{20} \end{bmatrix}$; $\frac{4}{5}, \frac{6}{5}$ [𝔸]-359

14. Justify the claim.

   (a) Similar matrices have the same trace (sum of the entries on the main diagonal).

   (b) Similar matrices have equal eigenspace dimensions.

Answers

reflection about ℓ The standard matrix for reflection about the x-axis followed by counterclockwise rotation by 2α about the origin is

$$\begin{bmatrix} \cos(2\alpha) & -\sin(2\alpha) \\ \sin(2\alpha) & \cos(2\alpha) \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = M.$$

inverse of P It is easiest to verify that the product of the two matrices is the identity:

$$\begin{aligned} \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}\begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} &= \begin{bmatrix} \cos^2\alpha + \sin^2\alpha & \cos\alpha\sin\alpha - \sin\alpha\cos\alpha \\ \sin\alpha\cos\alpha - \cos\alpha\sin\alpha & \cos^2\alpha + \sin^2\alpha \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

eigenvalues and eigenvectors The characteristic equation is

$$\begin{aligned} \begin{vmatrix} -7-\lambda & -4 \\ 6 & 7-\lambda \end{vmatrix} &= (-7-\lambda)(7-\lambda) - (-4)(6) \\ &= -49 + \lambda^2 + 24 \\ &= \lambda^2 - 25 = 0 \end{aligned}$$

so the eigenvalues are -5 and 5. Eigenvectors are found by reducing the matrices

$$\begin{bmatrix} -2 & -4 \\ 6 & 12 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -12 & -4 \\ 6 & 2 \end{bmatrix}$$

to

$$\begin{bmatrix} -2 & -4 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -12 & -4 \\ 0 & 0 \end{bmatrix},$$

which gives eigenvectors of the forms $x = -2y$ for $\lambda = -5$ and $y = -3x$ for $\lambda = 5$. Choosing eigenvectgors with small integer entries, we take eigenpairs

$$-5, \begin{bmatrix} -2 \\ 1 \end{bmatrix} \quad \text{and} \quad 5, \begin{bmatrix} 1 \\ -3 \end{bmatrix}.$$

coordinates with respect to C The change of coordinates matrix from the standard basis to C is $\begin{bmatrix} -\frac{3}{5} & -\frac{1}{5} \\ -\frac{1}{5} & -\frac{2}{5} \end{bmatrix}$, so point P has coordinates

$$\begin{bmatrix} -\frac{3}{5} & -\frac{1}{5} \\ -\frac{1}{5} & -\frac{2}{5} \end{bmatrix}\begin{bmatrix} -2 \\ 3 \end{bmatrix}_{\mathcal{E}} = \begin{bmatrix} \frac{3}{5} \\ -\frac{4}{5} \end{bmatrix}_{C}.$$

coordinates with respect to $\mathcal{C}$ As can be seen from figure 5.4.2, the coordinates of $S(P)$ are $\begin{bmatrix} 2 \\ 9 \end{bmatrix}_{\mathcal{E}}$, consistent with (5.4.3), which shows $S\left(\begin{bmatrix} -2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 9 \end{bmatrix}$.

Part III

Applications

Chapter 6

Mathematical Applications

6.1 LU Factorization [3.4, 3.6]

Humans and computers bring very different skills to task. Computers, as the name would suggest, are famously adept at computing. Handling fractions, square roots, irrational constants, and $1+2$ are all the same to a computer. Humans tend to balk at non-integer computations and lengthy ones too. Ask a human to execute five, six, or twenty numerical operations, even integer operations, to solve a single problem and they might think you were simply asking too much. Ask a computer to do the same and you'll have your answer in just a few milliseconds.

Computers supply speed and accuracy to what would be tedious and error prone for a human. Computers make the otherwise impractical practical. Linear systems with a hundred, thousand, or even a hundred thousand equations are easily within the practical limits of computers. Even a home computer can handle a few hundred equations. However, at that size, efficiency plays a major role in practicality. An efficient algorithm may be able to handle a certain computation in a few seconds while an inefficient one may take a few days for the same. Humans provide appropriate and fast algorithms that would, at least until AI makes some huge advances, be impossible for a computer.

What makes a good algorithm is first and foremost accuracy. If the algorithm does not conclude with the correct answer, it is of no use at all. What makes one correct algorithm better than another is speed, measured by comparing the number of computations needed to complete the computation with common functions such as n^2 or 3^n where n is some measure of the "size" of the problem. Such a function gives a good idea how the time it takes to solve a problem grows as the size grows. To get a sense of the computer's speed in solving a linear system, let n represent the number of equations to be solved. The row reduction algorithm of section 2.2, also known as Gaussian elimination, requires $\frac{n^3}{3} + n^2 - \frac{n}{3}$ multiplications/divisions and $\frac{n^3}{3} + \frac{n^2}{2} - \frac{5n}{6}$ additions/subtractions ([4] section 6.1) to execute. Table 6.1 lists the number of arithmetic operations required to compute reduced row echelon form for a general linear system with n equations in n unknowns. For a human with a handheld calculator, systems with 2 or 3 variables are doable ($n = 2$ or $n = 3$). Systems with 4 variables would be tedious in general, but practical with a few strategically placed zeros. However, somewhere between 5 and 10 equations we find the limit of human practicality for general systems.

Table 6.1: Arithmetic operations required for row reduction

equations, n	$\times/\div$	$+/-$	total	$\frac{2}{3}n^3$
2	6	3	9	$5+\frac{1}{3}$
3	17	11	28	18
4	36	26	62	$42+\frac{2}{3}$
9	321	276	597	486
51	46,801	45,475	92,276	88,434
102	364,106	358,853	722,959	707,472
501	42,168,001	42,042,250	84,210,251	83,834,334
2,001	2,674,672,001	2,672,669,000	5,347,341,001	5,341,337,334
10,002	333,633,410,006	333,583,385,003	667,216,795,009	667,066,746,672

You might well ask whether systems of 500 equations in 500 unknowns are practical even for a computer. Their solution by Gaussian elimination requires over 83 million computations! Even if the computer does one computation every microsecond (millionth of a second), solving a system of 500 equations in 500 unknowns will take about 83 seconds. Depending on how quickly results are needed, this may or may not be practical. The same computer would take over $5,300$ seconds (about an hour and a half) to solve a system of $2,000$ equations in $2,000$ unknowns and over $666,000$ seconds (over 7 and a half days!) to solve a system of $10,000$ equations in $10,000$ unknowns. Of course faster computers or clusters of computers could be put to the task to speed up the computation, but no matter how much computing power is supplied, there will always be a less-than-astronomical size outside its practical limits.

A better option than more computing power is to streamline the algorithm. The numerical "speed" of Gaussian elimination is approximately proportional to n^3. The rightmost column of table 6.1 illustrates this point. The number of computations can be well approximated by $\frac{2}{3}n^3$ for large n. In the parlance of numerical analysis, one would say the algorithm executes in $O(n^3)$ time, read "big-oh of n^3 time". The implication is doubling the size of the problem multiplies the time it takes to execute by 8 and more generally increasing the size by a factor of k increases its execution time by a factor of k^3.

If the algorithm executed in, say, $O(n^2)$ time it would reduce the number of computations needed to solve large problems by several orders of magnitude. For example, an algorithm that required approximately $4n^2$ computations would need "only" about $1,000,000$ computations to solve a system of 500 equations in 500 unknowns (compare this to the 83 million for Gausian elimination); about $16,000,000$ for a system of $2,000$ equations in $2,000$ unknowns (compare this to the over 53 bilion for Gaussian elimination); and about 400 million for a system of $10,000$ equations in $10,000$ unknowns (reducing the estimated time of 7.5 days to about 6 minutes!).

Now imagine you have to solve the system multiple times for multiple sets of constants, but the same coefficient matrix, something that happens frequently in industrial applications. 7.5 days is preferable to 75 days for solving the system with ten different sets of constants. This is the approximate effect and purpose of using LU factorization—solving large systems for multiple sets of constants. The factorization itself takes about the same effort as Gaussian elimination, but subsequent solutions take $O(n^2)$ time.

The efficiency gain is due to turning the general problem into a special case that is much quicker to solve. As noted earlier, asking a human with a handheld calculator to solve a system of 4 equations in 4 unknowns borders on the impractical as it requires as many as 62 individual arithmetic operations. But asking a human to solve a system of 4 equations in 4 unknowns where the coefficient matrix is upper triangular is well within reason. You might have a go at solving the system

$$\begin{bmatrix} -14 & 0 & -10 & 12 \\ 0 & 9 & -8 & 4 \\ 0 & 0 & -15 & -6 \\ 0 & 0 & 0 & 10 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -11 \\ -2 \\ 6 \\ 15 \end{bmatrix}$$

to see that it takes 26 arithmetic operations—still not terribly appealing but much better than the 62 for Gaussian elimination of a general 4-equation, 4-variable system. A similar reduction is achieved when the coefficient matrix is lower triangular. These observations are at the heart of the LU factorization (lower-upper factorization) algorithm. It requires many fewer computations to solve two linear systems, one with an upper triangular coefficient matrix and the other with a lower triangular coefficient matrix, than it does to solve a single general linear system. In general, solving a system with either an upper triangular or lower triangular coefficient matrix can be done in $O(n^2)$ time. The number of computations needed is approximately proportional to n^2.

To review, if we could factor a general coefficient matrix M into the product of a lower triangular and upper triangular matrix, we could achieve such efficiency. Solving $M\mathbf{v} = \mathbf{b}$ directly by Gaussian elimination requires $O(n^3)$ operations while solving $LU\mathbf{v} = \mathbf{b}$ "twice"—first to find $U\mathbf{v}$ by solving $L\mathbf{w} = \mathbf{b}$ and second to find $\mathbf{v}$ by solving $U\mathbf{v} = \mathbf{w}$—requires only $O(n^2)$ operations.

The process for factoring M into LU essentially amounts to saving your progress in executing Gaussian elimination to the point where M has first reached echelon form. As we did several times in sections 3.6 and 3.7, we will rely on the fact that row reduction can be expressed by multiplication by (invertible) elementary matrices. If we record the row operations (elementary matrices) that reduce M to echelon form, we have

$$E_p \cdots E_2 E_1 M = U$$

where U (the echelon form of M) is upper triangular. Assuming no row swaps have been done, the product $E_p \cdots E_2 E_1$ is lower triangular since all row replacements are done by adding a multiple of one row to a row below it (and row scaling does not affect the locations of zeros). By the same reasoning the inverse of $E_p \cdots E_2 E_1$ is lower triangular, so setting $L^{-1} = E_p \cdots E_2 E_1$ we have

$$M = (E_p \cdots E_2 E_1)^{-1} U = LU.$$

It is not realistic to expect that reduction can be done without row swaps, however. If, for example, the 1,1-entry of M is zero and there is at least one nonzero entry below it, there is no choice. The first operation of row reduction requires a row swap. This means M cannot always be factored into a lower triangular times upper triangular product, and allowing row swaps is necessary. In the end, a factorization that includes row swaps does not factor M into a product LU. Instead, it factors M into a product PLU where P is a **permutation matrix**, a matrix that holds the same rows as the identity matrix but in a possibly different order. Including P in the computation this way does not add any arithmetic operations to the algorithm. It adds only swapping of values, a computer operation that is so fast as to be insignificant compared to arithmetic operations. Allowing row swaps in an LU decomposition is known as LU decomposition with **partial pivoting**.

An Example

To illustrate the method, we will factor (factorize, or decompose)

$$M = \begin{bmatrix} 18 & -35 & -4 & -56 \\ -14 & 21 & 4 & 42 \\ 6 & -7 & -1 & -23 \\ 6 & -9 & -2 & -18 \end{bmatrix}.$$

operations	result
$M_{1,:} \leftrightarrow M_{4,:}$	$\begin{bmatrix} 6 & -9 & -2 & -18 \\ -14 & 21 & 4 & 42 \\ 6 & -7 & -1 & -23 \\ 18 & -35 & -4 & -56 \end{bmatrix}$
$M_{1,:} \to \frac{1}{3} M_{1,:}$	$\begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \\ -14 & 21 & 4 & 42 \\ 6 & -7 & -1 & -23 \\ 18 & -35 & -4 & -56 \end{bmatrix}$
$\begin{array}{ll} M_{2,:} & \to 7M_{1,:} + M_{2,:} \\ M_{3,:} & \to -3M_{1,:} + M_{3,:} \\ M_{4,:} & \to -9M_{1,:} + M_{4,:} \end{array}$	$\begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \\ 0 & 0 & -\frac{2}{3} & 0 \\ 0 & 2 & 1 & -5 \\ 0 & -8 & 2 & -2 \end{bmatrix}$
$M_{2,:} \leftrightarrow M_{3,:}$	$\begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \\ 0 & 2 & 1 & -5 \\ 0 & 0 & -\frac{2}{3} & 0 \\ 0 & -8 & 2 & -2 \end{bmatrix}$
$M_{4,:} \to 4M_{2,:} + M_{4,:}$	$\begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \\ 0 & 2 & 1 & -5 \\ 0 & 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 6 & -22 \end{bmatrix}$
$M_{4,:} \to 9M_{3,:} + M_{4,:}$	$\begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \\ 0 & 2 & 1 & -5 \\ 0 & 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & -22 \end{bmatrix}$

We have reached echelon form, so we have identified U (the echelon form itself):

$$U = \begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \\ 0 & 2 & 1 & -5 \\ 0 & 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & -22 \end{bmatrix}$$

and the product of the 8 elementary matrices corresponding to the 8 row operations form PL. Apply the inverse of each row operation, in reverse order, starting with the identity matrix:

$$\begin{array}{ll} \text{operation} & \text{inverse} \\ M_{4,:} \to 9M_{3,:} + M_{4,:} & M_{4,:} \to -9M_{3,:} + M_{4,:} \\ M_{4,:} \to 4M_{2,:} + M_{4,:} & M_{4,:} \to -4M_{2,:} + M_{4,:} \\ M_{2,:} \leftrightarrow M_{3,:} & M_{2,:} \leftrightarrow M_{3,:} \\ M_{4,:} \to -9M_{1,:} + M_{4,:} & M_{4,:} \to 9M_{1,:} + M_{4,:} \\ M_{3,:} \to -3M_{1,:} + M_{3,:} & M_{3,:} \to 3M_{1,:} + M_{3,:} \\ M_{2,:} \to 7M_{1,:} + M_{2,:} & M_{2,:} \to -7M_{1,:} + M_{2,:} \\ M_{1,:} \to (1/3)M_{1,:} & M_{1,:} \to 3M_{1,:} \\ M_{1,:} \leftrightarrow M_{4,:} & M_{1,:} \leftrightarrow M_{4,:} \end{array}$$

In full detail:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -9 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -4 & -9 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -4 & -9 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 9 & -4 & -9 & 1 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 3 & 1 & 0 & 0 \\ 9 & -4 & -9 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 \\ -7 & 0 & 1 & 0 \\ 3 & 1 & 0 & 0 \\ 9 & -4 & -9 & 1 \end{bmatrix} \to \begin{bmatrix} 3 & 0 & 0 & 0 \\ -7 & 0 & 1 & 0 \\ 3 & 1 & 0 & 0 \\ 9 & -4 & -9 & 1 \end{bmatrix} \to \begin{bmatrix} 9 & -4 & -9 & 1 \\ -7 & 0 & 1 & 0 \\ 3 & 1 & 0 & 0 \\ 3 & 0 & 0 & 0 \end{bmatrix} = PL$$

At this point, we have

$$M = \begin{bmatrix} 9 & -4 & -9 & 1 \\ -7 & 0 & 1 & 0 \\ 3 & 1 & 0 & 0 \\ 3 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \\ 0 & 2 & 1 & -5 \\ 0 & 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & -22 \end{bmatrix},$$

which, as expected, is not lower triangular times upper triangular. The final step is to permute the rows of M. The permutation matrix P^{-1} can be constructed by applying only the row swaps to the identity matrix, in the same order in which they were applied during row reduction. In this case, that means $M_{1,:} \leftrightarrow M_{4,:}$ and then $M_{2,:} \leftrightarrow M_{3,:}$:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} = P^{-1}.$$

In this case the order of the swaps does not matter, but when an index is repeated within the set of swaps, the order will matter. Finally, we have

$$P^{-1}M = LU = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -7 & 0 & 1 & 0 \\ 9 & -4 & -9 & 1 \end{bmatrix} \begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \\ 0 & 2 & 1 & -5 \\ 0 & 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & -22 \end{bmatrix}$$

or $M = PLU$. Can you verify this? Answer on page 200.

Therefore the system $M\mathbf{v} = \mathbf{b}$ is equivalent to $PLU\mathbf{v} = \mathbf{b}$. Solving amounts to first applying P^{-1} to $\mathbf{b}$, permuting its entries, yielding $LU\mathbf{v} = P^{-1}\mathbf{b}$; second, solving $L\mathbf{w} = P^{-1}\mathbf{b}$, which yields $\mathbf{w} = L^{-1}P^{-1}\mathbf{b}$; and third, solving $U\mathbf{v} = \mathbf{w}$, which yields $\mathbf{v}$ such that $U\mathbf{v} = L^{-1}P^{-1}\mathbf{b}$.

When M is invertible, U will be invertible, and we will have $\mathbf{v} = U^{-1}L^{-1}P\mathbf{b} = M^{-1}\mathbf{b}$, but the method works even when M is noninvertible. Whenever $U\mathbf{v} = \mathbf{w}$ is consistent, its solutions will also be solutions of $M\mathbf{v} = \mathbf{b}$ (and whenever $U\mathbf{v} = \mathbf{w}$ is inconsistent $M\mathbf{v} = \mathbf{b}$ will also be inconsistent).

Key Concepts

factorization factoring a matrix into a product of matrices.

decomposition factorization.

LU **factorization** factoring a matrix into a product of a lower triangular matrix (L) by an upper triangular matrix (U).

partial pivoting allowing row swaps in *LU* factorization. In this case, the algorithm factorizes PM into LU for some permutation matrix P.

permutation matrix a matrix containing the same rows as an identity matrix but in a possibly different order. Such a matrix will have exactly one 1 in each row and each column and zeros elsewhere.

LU **advantage** solving a system in $O(n^3)$ time, subsequently solving systems with the same coefficient matrix but different constants in $O(n^2)$ time.

Exercises

1. Provide an *LU* factorization of M (no row swaps during row reduction).

   (a) $M = \begin{bmatrix} 3 & 5 \\ -9 & -14 \end{bmatrix}$ [𝔸]-359

   (b) $M = \begin{bmatrix} 5 & -6 \\ -10 & 14 \end{bmatrix}$

   (c) $M = \begin{bmatrix} -2 & -6 \\ 5 & 35 \end{bmatrix}$ [𝔸]-359

   (d) $M = \begin{bmatrix} 21 & -7 \\ 3 & -1 \end{bmatrix}$

   (e) $M = \begin{bmatrix} 4 & 24 & 24 \\ 1 & -30 & 0 \end{bmatrix}$ [𝕊]-325

   (f) $M = \begin{bmatrix} -18 & -36 & -12 \\ 12 & 29 & -12 \end{bmatrix}$

   (g) $M = \begin{bmatrix} -25 & 10 \\ -20 & -13 \\ 5 & -8 \end{bmatrix}$ [𝔸]-359

   (h) $M = \begin{bmatrix} -42 & -7 \\ 24 & 16 \\ -24 & 26 \end{bmatrix}$

   (i) $M = \begin{bmatrix} -6 & 8 & 4 \\ 6 & 4 & -4 \\ -9 & -6 & 0 \end{bmatrix}$ [𝔸]-359

   (j) $M = \begin{bmatrix} 21 & -7 & -35 \\ -3 & 37 & 11 \\ -3 & -41 & -10 \end{bmatrix}$

2. Find a permutation matrix $P \neq I$ and an *LU* factorization of PM. Use partial pivoting (at least one row swap during row reduction).

   (a) $M = \begin{bmatrix} -4 & -3 \\ 1 & 7 \end{bmatrix}$ [𝔸]-359

   (b) $M = \begin{bmatrix} -10 & -9 \\ 1 & -1 \end{bmatrix}$

   (c) $M = \begin{bmatrix} -7 & 3 \\ 6 & -10 \\ 8 & -5 \end{bmatrix}$ [𝕊]-326

   (d) $M = \begin{bmatrix} 4 & -7 \\ 8 & -2 \\ -6 & 12 \end{bmatrix}$

   (e) $M = \begin{bmatrix} 8 & -10 & -4 \\ -1 & 2 & 0 \\ -12 & 12 & 1 \end{bmatrix}$ [𝔸]-359

   (f) $M = \begin{bmatrix} 10 & 1 & -10 \\ -7 & 3 & -4 \\ 5 & -1 & 7 \end{bmatrix}$

3. Use the *LU* decomposition to calculate the determinant of M in question 1. [𝕊]-326 [𝔸]-359

4. Use the *LU* decomposition to calculate the determinant of M in question 2. [𝕊]-326 [𝔸]-359

5. *LU* factorizations are not unique. Redo question 1 without using row swaps or row scaling, thus producing a factorization where L has 1s on its diagonal. [𝕊]-326 [𝔸]-359

6. *LU* factorizations are not unique. Redo question 1 (without using row swaps) but use row scaling so that all the nonzero diagonal entries of U are 1. [𝔸]-359

7. Use the fact that $M = LU$ to help solve the system $M\mathbf{v} = \mathbf{b}$.

   (a) $M = \begin{bmatrix} 6 & -6 \\ 42 & -35 \end{bmatrix}$; $L = \begin{bmatrix} 1 & 0 \\ 7 & 1 \end{bmatrix}$; $U = \begin{bmatrix} 6 & -6 \\ 0 & 7 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} -6 \\ -21 \end{bmatrix}$

   (b) $M = \begin{bmatrix} -5 & 1 \\ -20 & -2 \end{bmatrix}$; $L = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}$; $U = \begin{bmatrix} -5 & 1 \\ 0 & -6 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} -23 \\ -14 \end{bmatrix}$ [𝔸]-360

   (c) $M = \begin{bmatrix} 4 & 7 \\ 4 & 2 \end{bmatrix}$; $L = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$; $U = \begin{bmatrix} 4 & 7 \\ 0 & -5 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} -2 \\ 8 \end{bmatrix}$

(d) $M = \begin{bmatrix} 3 & -7 & -2 \\ -12 & 35 & 7 \\ -6 & 28 & 4 \end{bmatrix}$;

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

$U = \begin{bmatrix} 3 & -7 & -2 \\ 0 & 7 & -1 \\ 0 & 0 & 2 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 0 \\ 6 \\ 14 \end{bmatrix}$ [𝕊]-326

(e) $M = \begin{bmatrix} -7 & -3 & -5 \\ -7 & -6 & 0 \\ -21 & 6 & -39 \end{bmatrix}$;

$L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & -5 & 1 \end{bmatrix}$;

$U = \begin{bmatrix} -7 & -3 & -5 \\ 0 & -3 & 5 \\ 0 & 0 & 1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 14 \\ 10 \\ 63 \end{bmatrix}$

(f) $M = \begin{bmatrix} 2 & -7 & -4 \\ -4 & 7 & 8 \\ 4 & -63 & -6 \end{bmatrix}$;

$L = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & 7 & 1 \end{bmatrix}$;

$U = \begin{bmatrix} 2 & -7 & -4 \\ 0 & -7 & 0 \\ 0 & 0 & 2 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 5 \\ -17 \\ -43 \end{bmatrix}$

[𝔸]-360

8. Redo question 7 solving $M\mathbf{v} = \mathbf{b}$ directly (without using L or U) instead. Compare the labor involved in the two methods.

9. Find the inverse of the matrix M of question 7 by computing $U^{-1}L^{-1}$. [𝔸]-360

10. Suppose M is an $n \times n$ matrix with decomposition LU and L has 1s on its diagonal. Count precisely the number of arithmetic operations (additions, subtractions, multiplications, and divisions) needed to solve the system $M\mathbf{v} = \mathbf{b}$ using the LU decomposition. It will be a quadratic function of n.

11. Rank factorization. Suppose M is an $m \times n$ matrix of rank $k \le \min\{m, n\}$. Argue that there are matrices P, C, F where $PM = CF$; P is a permutation matrix; C is $m \times k$; and F is $k \times n$.

Answers

LU-factorization verified

$$P^{-1}M = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 18 & -35 & -4 & -56 \\ -14 & 21 & 4 & 42 \\ 6 & -7 & -1 & -23 \\ 6 & -9 & -2 & -18 \end{bmatrix} = \begin{bmatrix} 6 & -9 & -2 & -18 \\ 6 & -7 & -1 & -23 \\ -14 & 21 & 4 & 42 \\ 18 & -35 & -4 & -56 \end{bmatrix}$$

and

$$LU = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -7 & 0 & 1 & 0 \\ 9 & -4 & -9 & 1 \end{bmatrix} \begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \\ 0 & 2 & 1 & -5 \\ 0 & 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & -22 \end{bmatrix}$$

so

$$\begin{aligned}
(LU)_{1,:} &= 3\begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \end{bmatrix} = \begin{bmatrix} 6 & -9 & -2 & -18 \end{bmatrix} \\
(LU)_{2,:} &= 3\begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \end{bmatrix} + 1\begin{bmatrix} 0 & 2 & 1 & -5 \end{bmatrix} \\
&= \begin{bmatrix} 6 & -7 & -1 & -23 \end{bmatrix} \\
(LU)_{3,:} &= -7\begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \end{bmatrix} + 1\begin{bmatrix} 0 & 0 & -\frac{2}{3} & 0 \end{bmatrix} \\
&= \begin{bmatrix} -14 & 21 & 4 & 42 \end{bmatrix} \\
(LU)_{4,:} &= 9\begin{bmatrix} 2 & -3 & -\frac{2}{3} & -6 \end{bmatrix} - 4\begin{bmatrix} 0 & 2 & 1 & -5 \end{bmatrix} \\
&\quad - 9\begin{bmatrix} 0 & 0 & -\frac{2}{3} & 0 \end{bmatrix} + 1\begin{bmatrix} 0 & 0 & 0 & -22 \end{bmatrix} \\
&= \begin{bmatrix} 18 & -35 & -4 & -56 \end{bmatrix}
\end{aligned}$$

6.2 The Power Method [3.5]

Students typically learn about the quadratic formula in high school or earlier. It provides a formula for the solutions of the equation

$$ax^2 + bx + c = 0$$

(the roots of $f(x) = ax^2 + bx + c$) for any values a, b, and c. You may remember it as

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

The formula can also be written as

$$x = \frac{-b \pm \left(b^2 - 4ac\right)^{\frac{1}{2}}}{2a},$$

which combines the coefficients of the equation using addition, subtraction, multiplication, division, and rational exponentiation.

What you may never have seen are the more elaborate formulas for solving the general cubic,

$$a_3x^3 + a_2x^2 + a_1x + a_0 = 0$$

and the general quartic,

$$a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0.$$

Crumpet 25: Cubic and Quartic Formulas

The formula for solving the general cubic polynomial is most often credited to Gerolamo Cardano as he is the first known to have published the result, in 1545. However, history reveals that Niccolò Tartaglia knew of the solution before Cardano, and Scipione del Ferro, who died circa 1526, before him. In the same work where Cardano published the solution of the cubic, *Ars Magna*, he also published the solution of the quartic. History has been kinder to the original solver of the quartic, however, most often crediting the result to Lodovico Ferrari, a student of Cardano.

Though the formulas are much more involved than the quadratic formula, each one uses nothing more than addition, subtraction, multiplication, division, and radical exponentiation. As a result, any solution of any polynomial equation up to degree four can be written down explicitly and exactly.

A famous result of Galois theory is that there are unsolvable fifth degree polynomials (quintics), and unsolvable polynomials of all higher degrees as well. Their roots have no closed form expression using addition, subtraction, multiplication, division, and rational exponentiation of its coefficients. Their values cannot be written down exactly using traditional operations. The modest-looking $f(x) = x^5 - 10x + 2$ is a classic example. Its roots cannot be written down exactly and explicitly. However, being a fifth degree polynomial with real coefficients, it must have at least one real root. For large negative values of x, say -100 or -1000, $f(x)$ is negative and for large positive values of x, say 100 or 1000, $f(x)$ is positive. Somewhere in between, $f(x)$ must be zero (as required by the intermediate value theorem). Despite the impossibility of a formula, SageMath and other computer algebra systems can still find roots of any polynomial. But how can the roots of a 5^{th} degree polynomial be found if there is no formula?

As discussed, there is no exact, explicit formula. There are, however, many methods for approximating such roots, each one capable of determining the roots to arbitrary precision. Getting back to finding the roots of $f(x) = x^5 - 10x + 2$, if we set $x_0 = 2$ and let $x_{i+1} = x_i - \frac{f(x_i)}{5x_i^4 - 10}$, we can generate a sequence of approximations that get more and more accurate as we proceed. $x_1 = x_0 - \frac{f(x_0)}{5x_0^4 - 10} = 2 - \frac{f(2)}{5(2)^4 - 10} = \frac{9}{5}$ and $x_2 = x_1 - \frac{f(x_1)}{5x_1^4 - 1} = \frac{9}{5} - \frac{f\left(\frac{9}{5}\right)}{5\left(\frac{9}{5}\right)^4 - 1} \approx 1.731847$. Rounded to six decimal places each, the sequence $x_0, x_1, \ldots$ begins

$$2,\ 1.8,\ 1.731847,\ 1.724390,\ 1.724306$$

and 1.724306 is a root of f accurate to six decimal places.[1]

This is an example of an iterative routine. A starting value is given (2 in this case) and a recursive formula $\left(x_{i+1} = x_i - \frac{f(x_i)}{5x_i^4 - 10}\right)$ is applied to generate the rest of the list. The output of one iteration is input into the formula to calculate the next. As more and more iterations are calculated, the output approaches the desired quantity. There are a number of iterative routines for linear algebra problems too.

Let

$$M = \begin{bmatrix} 7 & 8 \\ -4 & -5 \end{bmatrix},$$

$\mathbf{v}_0 = \begin{bmatrix} -1 & 0 \end{bmatrix}^T$ and $\mathbf{v}_{i+1} = M\mathbf{v}_i$. With the initial value $\mathbf{v}_0$ and recursive formula, $\mathbf{v}_{i+1} = M\mathbf{v}_i$, we can compute a sequence of vectors:

$$\begin{aligned}
\mathbf{v}_1 &= M\mathbf{v}_0 = \begin{bmatrix} -7 & 4 \end{bmatrix}^T \\
\mathbf{v}_2 &= M\mathbf{v}_1 = \begin{bmatrix} -17 & 8 \end{bmatrix}^T \\
\mathbf{v}_3 &= M\mathbf{v}_2 = \begin{bmatrix} -55 & 28 \end{bmatrix}^T \\
\mathbf{v}_4 &= M\mathbf{v}_3 = \begin{bmatrix} -161 & 80 \end{bmatrix}^T \\
\mathbf{v}_5 &= M\mathbf{v}_4 = \begin{bmatrix} -487 & 244 \end{bmatrix}^T \\
\mathbf{v}_6 &= M\mathbf{v}_5 = \begin{bmatrix} -1457 & 728 \end{bmatrix}^T \\
\mathbf{v}_7 &= M\mathbf{v}_6 = \begin{bmatrix} -4375 & 2188 \end{bmatrix}^T \\
\mathbf{v}_8 &= M\mathbf{v}_7 = \begin{bmatrix} -13121 & 6560 \end{bmatrix}^T \\
\mathbf{v}_9 &= M\mathbf{v}_8 = \begin{bmatrix} -39367 & 19684 \end{bmatrix}^T \\
\mathbf{v}_{10} &= M\mathbf{v}_9 = \begin{bmatrix} -118097 & 59048 \end{bmatrix}^T
\end{aligned}$$

Though it is likely not apparent, something remarkable is happening here. Each vector is closer than the last to a vector of interest. Plotting the vectors helps reveal the phenomenon. Figure 6.2.1 shows eleven lines, one in the direction of each $\mathbf{v}_i$. The head of each $\mathbf{v}_i$ is marked with a point though it is difficult to see since $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_6$ are nearly on top of one another. Nonetheless, the figure illustrates what is happening. The slopes of the lines are converging (approaching a particular value). The last three lines, $\{r\mathbf{v}_8 : r \in \mathbb{R}\}$, $\{r\mathbf{v}_9 : r \in \mathbb{R}\}$, and $\{r\mathbf{v}_{10} : r \in \mathbb{R}\}$ all seem to lie more or less on the line $y = -\frac{1}{2}x$. That is, they essentially lie in the direction of $\begin{bmatrix} -2 & 1 \end{bmatrix}^T$. Further iteration will reveal more of the same.

What if $\mathbf{v}_0$ were different, though? Can you calculate a similar sequence of vectors starting with a different $\mathbf{v}_0$? Do the vectors of your sequence approach the direction $\begin{bmatrix} -2 & 1 \end{bmatrix}^T$ too? Answers with $\mathbf{v}_0 = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$ on page 208.

Figure 6.2.1 nicely illustrates the convergence geometrically, but we ought be able to detect it algebraically as

[1] You may recognize this as Newton's method.

Figure 6.2.1: Calculating $\mathbf{v}_{i+1} = M\mathbf{v}_i$ shows a sort of convergence

well. Let $\hat{\mathbf{v}}_i = \frac{1}{(\mathbf{v}_i)_{2,1}}\mathbf{v}_i$, $i = 1, 2, \ldots, 10$. That is, let $\hat{\mathbf{v}}_i$ be $\mathbf{v}_i$ scaled by the reciprocal of its second entry. Then

$$\begin{aligned}
\hat{\mathbf{v}}_1 &= \mathbf{v}_1/(\mathbf{v}_1)_{2,1} = \begin{bmatrix} -1.75 & 1 \end{bmatrix}^T \\
\hat{\mathbf{v}}_2 &= \mathbf{v}_2/(\mathbf{v}_2)_{2,1} = \begin{bmatrix} -2.125 & 1 \end{bmatrix}^T \\
\hat{\mathbf{v}}_3 &= \mathbf{v}_3/(\mathbf{v}_3)_{2,1} \approx \begin{bmatrix} -1.96429 & 1 \end{bmatrix}^T \\
\hat{\mathbf{v}}_4 &= \mathbf{v}_4/(\mathbf{v}_4)_{2,1} = \begin{bmatrix} -2.0125 & 1 \end{bmatrix}^T \\
\hat{\mathbf{v}}_5 &= \mathbf{v}_5/(\mathbf{v}_5)_{2,1} \approx \begin{bmatrix} -1.99590 & 1 \end{bmatrix}^T \\
\hat{\mathbf{v}}_6 &= \mathbf{v}_6/(\mathbf{v}_6)_{2,1} \approx \begin{bmatrix} -2.00137 & 1 \end{bmatrix}^T \\
\hat{\mathbf{v}}_7 &= \mathbf{v}_7/(\mathbf{v}_7)_{2,1} \approx \begin{bmatrix} -1.99954 & 1 \end{bmatrix}^T \\
\hat{\mathbf{v}}_8 &= \mathbf{v}_8/(\mathbf{v}_8)_{2,1} \approx \begin{bmatrix} -2.00015 & 1 \end{bmatrix}^T \\
\hat{\mathbf{v}}_9 &= \mathbf{v}_9/(\mathbf{v}_9)_{2,1} \approx \begin{bmatrix} -1.99995 & 1 \end{bmatrix}^T \\
\hat{\mathbf{v}}_{10} &= \mathbf{v}_{10}/(\mathbf{v}_{10})_{2,1} \approx \begin{bmatrix} -2.00002 & 1 \end{bmatrix}^T
\end{aligned}$$

Being a scalar multiple, the vector $\hat{\mathbf{v}}_i$ points in the same direction as $\mathbf{v}_i$. The list of $\hat{\mathbf{v}}_i$ numerically demonstrates that the $\hat{\mathbf{v}}_i$, and therefore the $\mathbf{v}_i$, are pointing closer and closer to the $\begin{bmatrix} -2 & 1 \end{bmatrix}^T$ direction. Interestingly,

$$M\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 & 8 \\ -4 & -5 \end{bmatrix}\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -6 \\ 3 \end{bmatrix} = 3\begin{bmatrix} -2 \\ 1 \end{bmatrix},$$

or $M\begin{bmatrix} -2 & 1 \end{bmatrix}^T = 3\begin{bmatrix} -2 & 1 \end{bmatrix}^T$, so $3, \begin{bmatrix} -2 & 1 \end{bmatrix}^T$ is an eigenpair of M!

Sit back and think about this for a moment. We started with a seemingly arbitrary matrix and a pretty shabby approximation of one of its eigenvectors. We then proceeded to multiply the (shabby) approximation by M, the resulting product by M again, the resulting product by M again, and so on, the final result of which was a very good approximation of an eigenvector of M. With a computer at hand to do the calculation, this is a whole lot easier than solving a characteristic equation. More importantly, though, remember general polynomials of degree five or higher cannot be solved exactly. This means eigenpairs for square matrices of size 5×5 and up cannot generally be found exactly (their characteristic polynomials are fifth degree or higher). Numerical methods must be used!

The approach of iteratively multiplying some vector by the matrix whose eigenpairs are desired, known as the **power method**, seems to have potential, but the example should leave you with lots of questions. For example,

Does this always work?
Why does it work?
If it doesn't always work, when does it work?
Are there other methods we can try?
Can we say how many iterations are needed to get a good approximation?
Which eigenpair is found when it works?
Does it matter what vector is chosen for $\mathbf{v}_0$?
Will the method always produce the same eigenvector?
What about other eigenvectors?

The computation in crumpet 26 answers several questions, partially answers others, and leaves some open. For example, the method works when

(i) M is $n \times n$ and has n linearly independent eigenvectors, and

(ii) one of the eigenvalues is **dominant** in the sense that its magnitude is larger than all the others, and

(iii) the eigenspace of the dominant eigenvalue has dimension one, and

(iv) $\mathbf{v}_0$ is chosen appropriately.

That's not to say it won't work in other instances. As mathematicians say, these are sufficient conditions, not necessary conditions. Other answers include

1. The method only determines the eigenvector corresponding to the dominant eigenvalue.

2. The accuracy of the approximation is proportional to the largest ratio $\left|\frac{\lambda_d}{\lambda_i}\right|^k$, $i \neq d$ where λ_d is the dominant eigenvalue.

The exercises explore a number of these points, and describes a modification of the power method that can be used to find any eigenpair.

On a practical note, implementation of the method will include scaling $\mathbf{v}_k$ with each iteration. As the example shows, the norm of $\mathbf{v}_k$ can grow very large very quickly. Crumpet 1 reveals that $\mathbf{v}_k$ will grow (or decay) exponentially. Since it is only the direction of $\mathbf{v}_k$ that matters, scaling does not affect the success of the algorithm. Typically, $\mathbf{v}_k$ will be scaled by

$$\max\{|(\mathbf{v}_k)_{j,1}| : j = 1, 2, \ldots, n\} \tag{6.2.1}$$

so that the magnitude of the greatest (or least) entry of $\mathbf{v}_k$ is one.

Crumpet 26: The Power Method

Suppose M is a diagonalizable $n \times n$ matrix and P is an $n \times n$ matrix whose columns are linearly independent eigenvectors of M (see section 5.4), and let $D = P^{-1}MP$. Further suppose that the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of M are such that for some d, λ_d is the dominant eigenvalue ($|\lambda_j| > |\lambda_i|$ for all $i \neq j$) and the eigenspace of λ_d has dimension one. Pick an arbitrary vector $\mathbf{v}_0$ in $\mathbb{R}^n$ and let $\mathbf{w} = P^{-1}\mathbf{v}_0$. Finally, define $\mathbf{v}_k = M\mathbf{v}_{k-1}$ for $k > 0$ and Z as the $n \times n$ matrix with zeros everywhere except the d,d-entry where it has a one. Then for large enough k,

$$\begin{aligned}
\mathbf{v}_k = M^k\mathbf{v}_0 &= PD^kP^{-1}\mathbf{v}_0 \\
&= PD^k\mathbf{w} \\
&= P\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}^k \mathbf{w} \\
&= P\begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{bmatrix} \mathbf{w} \\
&= \lambda_d^k P\begin{bmatrix} \frac{\lambda_1^k}{\lambda_d^k} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2^k}{\lambda_d^k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_n^k}{\lambda_d^k} \end{bmatrix} \mathbf{w} \\
&\approx \lambda_d^k PZ\mathbf{w} \\
&= \lambda_d^k \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & P_{:,d} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \mathbf{w} \\
&= \lambda_d^k \mathbf{w}_{d,1} P_{:,d}.
\end{aligned}$$

But $P_{:,d}$ is an eigenvector of M corresponding to λ_d, so $\mathbf{v}_k$ is approximately an eigenvector of M corresponding to λ_d as long as $\mathbf{w}_{d,1} \neq 0$.

Key Concepts

dominant eigenvalue an eigenvalue with larger magnitude than all other eigenvalues of a given matrix.

power method iteration of the recurrence relation $\mathbf{v}_k = M\mathbf{v}_{k-1}$ for some initial vector $\mathbf{v}_0$, usually with scaling. Under certain conditions the sequence $\mathbf{v}_0, \mathbf{v}_1, \ldots$ will converge to an eigenvector of M.

Exercises

1. Find the dominant eigenvalue of M.

   (a) $M = \begin{bmatrix} 11 & -4 \\ -6 & 9 \end{bmatrix}$

   (b) $M = \begin{bmatrix} -4 & 2 \\ -3 & -9 \end{bmatrix}$ [$\mathbb{S}$]-327

   (c) $M = \begin{bmatrix} 5 & 2 \\ 1 & 4 \end{bmatrix}$

   (d) $M = \begin{bmatrix} 5 & 1 \\ 2 & 6 \end{bmatrix}$ [$\mathbb{A}$]-360

   (e) $M = \begin{bmatrix} -27 & -29 & 39 \\ 30 & 32 & -45 \\ 4 & 4 & -7 \end{bmatrix}$

   (f) $M = \begin{bmatrix} -25 & 129 & -72 \\ 1 & -17 & 8 \\ 9 & -69 & 36 \end{bmatrix}$ [$\mathbb{A}$]-360

   (g) SageMathCell 68
   $M = \begin{bmatrix} 185 & 76 & 9 & 6 \\ -573 & -234 & -27 & -18 \\ 723 & 288 & 31 & 18 \\ 417 & 166 & 15 & 14 \end{bmatrix}$

   (h) SageMathCell 69
   $M = \begin{bmatrix} -8 & -72 & 72 & -0 \\ -39 & -110 & 88 & 3 \\ -39 & -126 & 104 & 3 \\ 2 & 108 & -104 & -18 \end{bmatrix}$
   [$\mathbb{A}$]-360

2. Does the matrix have a dominant eigenvalue?

(a) $\begin{bmatrix} -3 & 5 \\ -1 & 3 \end{bmatrix}$

(b) $\begin{bmatrix} 19 & 12 \\ -28 & -19 \end{bmatrix}$ [𝕊]-327

(c) SageMathCell 70 $\begin{bmatrix} 641 & 225 & -530 \\ -735 & -279 & 630 \\ 360 & 120 & -396 \end{bmatrix}$

(d) SageMathCell 71 $\begin{bmatrix} -35 & -5 & 10 \\ -137 & 1 & 58 \\ -76 & 8 & 14 \end{bmatrix}$
[𝔸]-360

(e) SageMathCell 72 $\begin{bmatrix} -6 & -11 & -32 \\ 72 & -68 & -128 \\ -18 & 11 & 8 \end{bmatrix}$

(f) SageMathCell 73
$\begin{bmatrix} -645 & 430 & -473 & 129 & -387 \\ -556 & 466 & -412 & -0 & -144 \\ -511 & 662 & -443 & -129 & 147 \\ -517 & 1106 & 209 & -817 & 1209 \\ 211 & -134 & 277 & -129 & 149 \end{bmatrix}$
[𝔸]-360

3. The sixth and seventh terms of the sequence defined by an initial vector, $\mathbf{v}_0$, and the recurrence $\mathbf{v}_k = M\mathbf{v}_{k-1}$, where M has a dominant eigenvalue, are given. Use this information to estimate an eigenpair of M.

(a) $\mathbf{v}_5 = \begin{bmatrix} -467 \\ 742 \end{bmatrix}$; $\mathbf{v}_6 = \begin{bmatrix} 1894 \\ -3020 \end{bmatrix}$
[𝕊]-327

(b) $\mathbf{v}_5 = \begin{bmatrix} 9311 \\ -3061 \end{bmatrix}$; $\mathbf{v}_6 = \begin{bmatrix} 93494 \\ -30994 \end{bmatrix}$

(c) $\mathbf{v}_5 = \begin{bmatrix} 4099 \\ 2049 \end{bmatrix}$; $\mathbf{v}_6 = \begin{bmatrix} 16381 \\ 8191 \end{bmatrix}$ [𝔸]-360

(d) $\mathbf{v}_5 = \begin{bmatrix} -27787 \\ 46793 \\ -18613 \end{bmatrix}$; $\mathbf{v}_6 = \begin{bmatrix} 282494 \\ -468970 \\ 187994 \end{bmatrix}$

(e) $\mathbf{v}_5 = \begin{bmatrix} 128272 \\ 177760 \\ -52800 \end{bmatrix}$; $\mathbf{v}_6 = \begin{bmatrix} 10654688 \\ 16132160 \\ -10084800 \end{bmatrix}$ [𝔸]-360

(f) $\mathbf{v}_5 = \begin{bmatrix} -9793 \\ -9793 \\ 7499 \end{bmatrix}$; $\mathbf{v}_6 = \begin{bmatrix} -1151951 \\ -1151951 \\ 546193 \end{bmatrix}$

4. Calculate $\mathbf{v}_1$ through $\mathbf{v}_{11}$ of the power method. Does it seem the method will converge?

(a) SageMathCell 74 $M = \begin{bmatrix} 23 & 4 \\ 30 & 21 \end{bmatrix}$;
$\mathbf{v}_0 = \begin{bmatrix} 11^{-8} \\ 11^{-8} \end{bmatrix}$ [𝕊]-328

(b) SageMathCell 75 $M = \begin{bmatrix} 2 & 1 \\ 4 & -1 \end{bmatrix}$;
$\mathbf{v}_0 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$

(c) SageMathCell 76 $M = \begin{bmatrix} 17 & 30 \\ -8 & -17 \end{bmatrix}$;
$\mathbf{v}_0 = \begin{bmatrix} 7^{-8} \\ 0 \end{bmatrix}$ [𝔸]-360

(d) SageMathCell 77 $M = \begin{bmatrix} -4 & -3 \\ 5 & 4 \end{bmatrix}$;
$\mathbf{v}_0 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

(e) SageMathCell 78 $M = \begin{bmatrix} -21 & 20 & -10 \\ -30 & 41 & -25 \\ -24 & 40 & -29 \end{bmatrix}$;
$\mathbf{v}_0 = \begin{bmatrix} 9^{-8} \\ 9^{-8} \\ 9^{-8} \end{bmatrix}$ [𝔸]-360

(f) SageMathCell 79 $M = \begin{bmatrix} 11 & -4 & 10 \\ 6 & 1 & 10 \\ -3 & 2 & 0 \end{bmatrix}$;
$\mathbf{v}_0 = \begin{bmatrix} \frac{1}{5} \\ -\frac{1}{5} \\ \frac{1}{5} \end{bmatrix}$

(g) SageMathCell 80 $M = \begin{bmatrix} -63 & 7 & 20 \\ -16 & -56 & 100 \\ -5 & 2 & -37 \end{bmatrix}$;
$\mathbf{v}_0 = \begin{bmatrix} 0 \\ 13^{-9} \\ 0 \end{bmatrix}$ [𝔸]-360

(h) SageMathCell 81 $M = \begin{bmatrix} -4 & 2 & 2 \\ 4 & 3 & -1 \\ -8 & 2 & 6 \end{bmatrix}$; $\mathbf{v}_0 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

5. Calculate $\mathbf{v}_6$ through $\mathbf{v}_{100}$ for the matrix and vector $\mathbf{v}_5$ of question 3, implementing the scaling procedure suggested in (6.2.1). Use this information to find an eigenpair of M exactly.

(a) SageMathCell 82 [𝕊]-328

(b) SageMathCell 83

(c) SageMathCell 84 [𝔸]-360

(d) SageMathCell 85

(e) SageMathCell 86 [𝔸]-360

(f) SageMathCell 87

6. The algebraic multiplicity of an eigenvalue is its multiplicity as a root of the characteristic polynomial. The geometric multiplicity of an eigenvalue is the dimension of its eigenspace. They are equal when the algebraic multiplicity is one. The geometric multiplicity is always less than or equal to the algebraic multiplicity. Given are the matrix M, its characteristic polynomial p, and the geometric multiplicity, g, of the dominant eigenvalue. (i) State the dominant eigenvalue of M. (ii) State the algebraic multiplicity of the dominant eigenvalue of M. (iii)

Apply the power method. Does it produce an (approximate) eigenvector? (iv) Apply the power method with a different initial vector. Does it produce an (approximate) eigenvector? Is this the same result as before?

(a) 88

$$M = \begin{bmatrix} -22 & -48 & -90 & -99 \\ -32 & -56 & 47 & 46 \\ -50 & -40 & -56 & -35 \\ 72 & 12 & -63 & -94 \end{bmatrix};$$

$p(\lambda) = (\lambda + 76)^2(\lambda + 57)(\lambda + 19)$; $g = 2$

(b) 89

$$M = \begin{bmatrix} 111 & 169 & -39 & -125 \\ -30 & -47 & 6 & 22 \\ 195 & 310 & -72 & -230 \\ 3 & -1 & -6 & -16 \end{bmatrix};$$

$p(\lambda) = (\lambda + 12)^2(\lambda + 3)(\lambda - 3)$; $g = 1$

[𝔸]-360

7. M is not diagonalizable (does not admit 4 linearly independent eigenvectors—see section 5.4) but has a dominant eigenvalue with only one associated eigenvector. Apply the power method anyway. Does it work?

(a) 90

$$M = \begin{bmatrix} -136 & 240 & -448 & 352 \\ 88 & -112 & 256 & -208 \\ 224 & -352 & 672 & -512 \\ 139 & -226 & 408 & -296 \end{bmatrix}$$

(b) 91

$$M = \begin{bmatrix} 14 & -134 & -34 & 59 \\ 18 & -114 & -14 & 55 \\ 48 & -512 & -68 & 248 \\ 48 & -352 & -40 & 172 \end{bmatrix} \quad [𝔸]\text{-}360$$

8. The dominant eigenvalue of M has algebraic multiplicity 2 but geometric multiplicity 1. See exercise 6 for an explanation of algebraic and geometric multiplicities.

$$M = \begin{bmatrix} 1528 & 876 & 736 & -436 \\ -1530 & -870 & -1080 & 360 \\ -431 & -267 & 124 & 197 \\ 782 & 618 & 500 & 76 \end{bmatrix}$$

The entrywise quotient of M^{65537} by M^{65536} is (approximately)

$$\begin{bmatrix} 390 & 390 & 390 & 390 \\ 390 & 390 & 390 & 390 \\ 390 & 390 & 390 & 390 \\ 390 & 390 & 390 & 390 \end{bmatrix}.$$

[𝔸]-360

(a) What does this say about the eigenvalues of M?

(b) What does this say about the eigenvectors of M? HINT: think about the columns of M^{65536}.

(c) What does this say about the power method applied to M?

9. 92 The eigenvalues of M are $-4, -20, -28$, and 28 so M does not have a dominant eigenvalue. It has two different eigenvalues with maximum magnitude. Running the power method with scaling as in (6.2.1) with the given $\mathbf{v}_0$ anyway gives $\mathbf{v}_{100}$ as shown. Express $\mathbf{v}_{100}$ as a linear combination of the eigenvectors $\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \\ -4 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 5 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ -5 \\ 3 \\ 0 \end{bmatrix}$ (corresponding to the eigenvalues $-4, -20, -28, 28$ in that order) of M.

$$M = \begin{bmatrix} 12 & 17 & 39 & -55 \\ 0 & 55 & 45 & -45 \\ 40 & 13 & 23 & -67 \\ 40 & 58 & 70 & -114 \end{bmatrix};$$

$$\mathbf{v}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}; \mathbf{v}_{100} \approx \begin{bmatrix} -0.477419 \\ 0.483871 \\ -1 \\ -0.709677 \end{bmatrix}$$

What does this say about $\mathbf{v}_{100}$. What does this say about the power method applied to M?

10. Find the eigenvalues of M from question 4. Use the information to supply an explanation of why the power method did/did not seem to converge. [𝔸]-360

Exercises 11-14 describe the inverse power method and supply one example.

11. Show that if $\lambda, \mathbf{v}$ is an eigenpair of M and M is invertible, then $\frac{1}{\lambda}, \mathbf{v}$ is an eigenpair of M^{-1}.

12. Show that if $\lambda, \mathbf{v}$ is an eigenpair of M then $\lambda - \alpha, \mathbf{v}$ is an eigenpair of $M - \alpha I$.

13. Combine 11 and 12 to show that if $\lambda, \mathbf{v}$ is an eigenpair of M and α is not an eigenvalue of M, then $\frac{1}{\lambda - \alpha}, \mathbf{v}$ is an eigenpair of $(M - \alpha I)^{-1}$.

14. 93 The power method could be used to find approximations of the dominant eigenvalue (approximately 77) and associated eigenvector of

$$M = \begin{bmatrix} 585 & -53 & -303 & -827 \\ 221 & 39 & -176 & -263 \\ 1652 & -36 & -944 & -2204 \\ -296 & -24 & 192 & 360 \end{bmatrix}.$$

With some preparation, however, the power method can be used to approximate non-dominant eigenpairs too! One of its eigenvalues, λ, is around 20. In fact 20 is closer to this particular eigenvalue than any of the others. Thus $\frac{1}{\lambda - 20}$ is the dominant eigenvalue of $(M - 20I)^{-1}$. Apply the power method to $(M - 20I)^{-1}$ to find approximations of λ and an associated eigenvector.

Answers

different $\mathbf{v}_0$ Setting $\mathbf{v}_0 = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$, for example, leads to the sequence $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{10}$

$$\begin{bmatrix} 15 & -9 \end{bmatrix}^T, \begin{bmatrix} 33 & -15 \end{bmatrix}^T, \begin{bmatrix} 111 & -57 \end{bmatrix}^T, \begin{bmatrix} 321 & -159 \end{bmatrix}^T, \begin{bmatrix} 975 & -489 \end{bmatrix}^T,$$
$$\begin{bmatrix} 2913 & -1455 \end{bmatrix}^T, \begin{bmatrix} 8751 & -4377 \end{bmatrix}^T, \begin{bmatrix} 26241 & -13119 \end{bmatrix}^T,$$
$$\begin{bmatrix} 78735 & -39369 \end{bmatrix}^T, \begin{bmatrix} 236193 & -118095 \end{bmatrix}^T.$$

Again the second entries are approximately $-\frac{1}{2}$ times the first, and getting closer the further we go in the sequence. For example, $\frac{-118095}{236193} \approx -0.499994$. Unless you chose $\mathbf{v}_0$ in the direction of $\begin{bmatrix} 1 & -1 \end{bmatrix}^T$, you should have noticed the same thing for your sequence.

Figure 6.3.1: What is the area of the overlapping region?

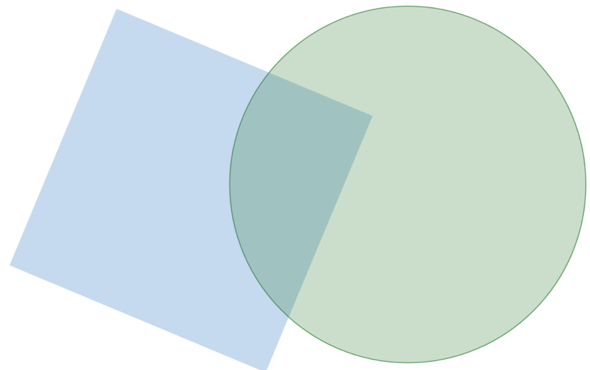
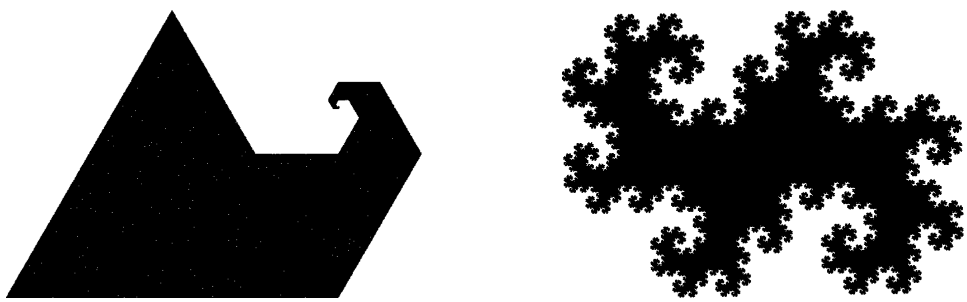


Figure 6.3.2: What are the areas of these shapes?



6.3 Geometry: Determinants, Eigenvalues, and Area [3.3, 3.6, 4.4]

Intuitively, we might think of area as the amount of paint needed to paint a particular shape. The more paint needed, the larger its area, and the larger its area, the more paint needed. To have some sense of what is meant by the area of an object, this intuition is good enough. Larger shapes have larger area while smaller shapes have smaller area, and the area of a shape is some measure of this size.

Calculating the areas of shapes (assigning numbers to areas) is another story. We certainly are not going to require that to find the area of an object it needs to be painted and the amount of paint used measured. What paint should be used, by whom, and what instrument should do the measuring? This process would be so imprecise as to be useless, giving the area of a single object many numerical areas. A single shape has but a single size, however, and so it must have but a single measure of its size—like the area formulas presented in grammar school. The area of a shape must be uniquely determined.

The area of a rectangle is its length times width. The area of a triangle is one half its base times height. The area of a circle is π times the square of its radius. Trapezoids, parallelograms, regular polygons, and unions of such shapes have calculable areas. But what about more complex shapes? For example, take an arbitrary nonempty overlap between a square and circle where neither is the circle contained within the square nor is the square contained within the circle. See figure 6.3.1. Calculus provides a method for calculating its area and hints at the complexity of the general question. By slicing the shape into smaller and smaller approximating rectangles and adding up the areas of those rectangles, the area can be approximated more and more accurately. The limit of these areas as the widths of the approximating rectangles approach zero is the area of the overlap. If you've taken calculus, that probably reminds you of integration, and it should! If you have not taken calculus, that probably sounds rather confusing and complicated, and it should! That is really the point. It is not an easy matter to calculate area, even of shapes that are easy to draw.

To stretch the point just a bit further, consider the shapes in figure 6.3.2. The figure on the left is the snail of Solomon Golomb[10] and features an infinitely spiraling appendage. The figure on the right is referred to as a twin dragon as it is the union of a pair of dragon curves. Neither of these figures can be drawn with perfect precision since each has infinitely small detail. The twin dragon is an example of a self-similar fractal with nonzero area. Its boundary (perimeter) is infinitely long and infinitely intricate. The more one magnifies the boundary, the more detail is revealed. While the snail can be formed by a union of infinitely many nonoverlapping triangles in a straightforward way, making its area calculable, the twin dragon cannot. Even applying calculus to the problem of finding the area of the twin dragon is not a straightforward matter. Does it even have a calculable area? What does having a calculable area mean? Are there sets whose areas are not calculable? These questions can be followed deep into measure theory, a branch of analysis far outside the reaches of this textbook.

With the very definition of area left as an interesting yet unresolved conundrum,

Crumpet 27: A Definition of Area

The area of a bounded region of the plane, a shape S, can be defined as follows. Let R be a polygonal region containing S, and let $\mathcal{P}_R$ be a primitive partition of R (a finite set of parallelograms and triangles whose interiors do not overlap and whose union is R). Define the norm of a partition, denoted $\|\mathcal{P}_R\|$, as the maximum of the areas of the primitives in $\mathcal{P}_R$. Then

$$\text{area}(S) = \lim_{\|\mathcal{P}_R\|\to 0} \sum_{\substack{p\in\mathcal{P}_R \\ p\subseteq S}} \text{area}(p)$$

whenever such limit exists.

it hardly makes practical sense to expect to prove the ways linear transformations affect the areas of general shapes. The following discussion is inherently incomplete this way. Certain claims regarding area will necessarily remain unproven.

Areas and determinants

In general, the image of a set S is defined as the set of images of all the points in S. That is, if S is a subset of A and $T : A \to B$, then the image of S under T is defined by $T(S) = \{T(s) : s \in S\}$. This definition is typical in all of mathematics, not just linear algebra, and applies no matter the sets A and B.

To understand how the linear transformation $T_A : \mathbb{R}^2 \to \mathbb{R}^2$, $T_A(\mathbf{v}) = A\mathbf{v}$ affects areas, it is convenient to write A as a product of elementary matrices, $A = E_p \cdots E_2E_1$, as we have done before, assuming A is invertible (page 106). Since $T_A(S) = (T_{E_p} \circ \cdots \circ T_{E_2} \circ T_{E_1})(S)$, if we can understand how linear transformations associated with elementary matrices affect area, we have a chance of understanding how general linear transformations affect area.

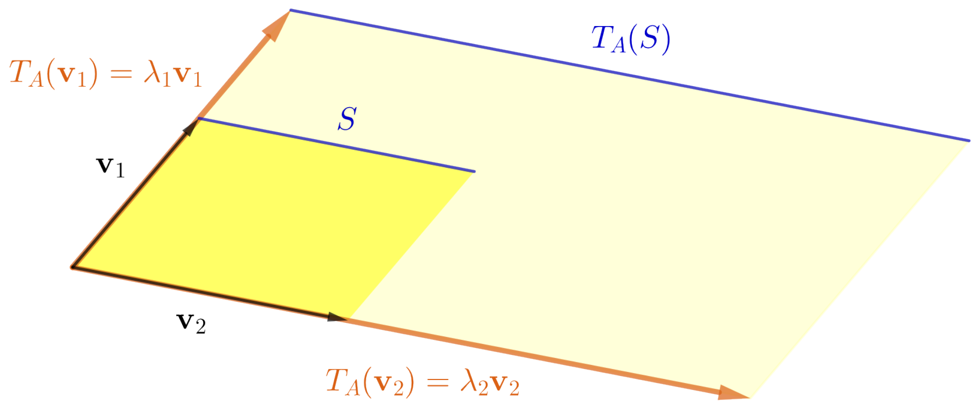
If E is a row swap matrix, then T_E is a reflection about the line $y = x$, so in this case $\text{area}(T_E(S)) = \text{area}(S)$. Reflections do not change areas. If E is a row replace matrix, then T_E is a shear transformation, and it is a known result of calculus that shear transformations do not affect area, so again $\text{area}(T_E(S)) = \text{area}(S)$. If E is a row scale matrix, then T_E scales shapes either horizontally or vertically—not both!—by a factor of s, so $\text{area}(T_E(S)) = |s| \cdot \text{area}(S)$. In every case, $\text{area}(T_E(S)) = |\det E| \cdot \text{area}(S)$ (the determinant of a row swap matrix is -1, the determinant of a row replace matrix is 1 and the determinant of a row scale matrix with scale factor s is s). It follows that

$$\begin{aligned}
\text{area}\,(T_A(S)) &= \text{area}\left((T_{E_p} \circ \cdots \circ T_{E_2} \circ T_{E_1})(S)\right) \\
&= \text{area}\left(T_{E_p}\,(\cdots(T_{E_2}\,(T_{E_1}(S)))\cdots)\right) \\
&= |\det E_p| \cdots |\det E_2| \cdot |\det E_1| \cdot \text{area}(S) \\
&= |\det A| \cdot \text{area}(S).
\end{aligned}$$

If A is noninvertible, then one of the columns of A is a multiple of the other, so any linear combination of the columns is also a multiple of that column. Therefore, the image of any vector, which is a linear combination of the columns of A, is a multiple of that column. Thus the image of every vector lies on the line determined by that column, giving the image of any shape zero area. The entire image is contained within a line. Of course, $|\det A| = 0$, so again we have $\text{area}(T_A(S)) = |\det A| \cdot \text{area}(S)$.

Areas and eigenvalues

Let A be a 2×2 matrix with linearly independent eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ corresponding to eigenvalues λ_1 and λ_2 respectively. Then $T_A(\mathbf{v}_1) = \lambda_1 \mathbf{v}_1$ and $T_A(\mathbf{v}_2) = \lambda_2 \mathbf{v}_2$. In fact, if we let $S = \{\mathbf{v}_1 + \alpha \mathbf{v}_2 : 0 \leq \alpha \leq 1\}$, the line segment from $\mathbf{v}_1$ to $\mathbf{v}_1 + \mathbf{v}_2$, then $T_A(S) = \{T_A(\mathbf{v}_1 + \alpha \mathbf{v}_2) : 0 \leq \alpha \leq 1\} = \{\lambda_1 \mathbf{v}_1 + \alpha \lambda_2 \mathbf{v}_2 : 0 \leq \alpha \leq 1\}$ is the line segment from $T_A(\mathbf{v}_1)$ to $T_A(\mathbf{v}_2)$. Further analysis of line segments shows that the image of the parallelogram determined by $\mathbf{v}_1$ and $\mathbf{v}_2$ is the parallelogram determined by $T_A(\mathbf{v}_1)$ and $T_A(\mathbf{v}_2)$. Can you supply this analysis? Answer on page 215.



Letting P be the parallelogram determined by $\mathbf{v}_1$ and $\mathbf{v}_2$, we see that T_A scales P in the $\mathbf{v}_1$ direction by a factor of λ_1 and in the $\mathbf{v}_2$ direction by factor λ_2. Therefore, the area of $T_A(P)$ equals $|\lambda_1 \lambda_2|$ times the area of P. Since we have been arguing that linear transformations scale the areas of all shapes the same way, we have generally that $\text{area}(T_A(S)) = |\lambda_1 \lambda_2| \cdot \text{area}(S)$ for any shape whose area is measurable. With respect to the eigenvectors of A, T_A is a simple scaling.

Now we have that

$$\text{area}(T_A(S)) = |\det A| \cdot \text{area}(S)$$
$$\text{and}$$
$$\text{area}(T_A(S)) = |\lambda_1 \lambda_2| \cdot \text{area}(S).$$

It must be, then, that $|\det A| = |\lambda_1 \lambda_2|$, a true statement about any 2×2 matrix! The statement can be made much stronger, however, as in the following theorem.

Theorem 15. [Determinant and the Product of Eigenvalues] *If A is an $n \times n$ matrix and $\lambda_1, \lambda_2, \ldots, \lambda_n$ are its n (possibly complex) eigenvalues, then*

$$\det A = \prod_{i=1}^{n} \lambda_i = \lambda_1 \lambda_2 \cdots \lambda_n.$$

Some but not all parts of the justification of this theorem are straightforward. For example, if A is upper triangular, then the conclusion follows quickly. As we have seen, $\det A$ is the product of the entries on the main diagonal. That is, $\det A = \prod_{i=1}^{n} A_{i,i}$. The characteristic equation

$$\begin{aligned} 0 &= \det(A - \lambda I) \\ &= \det\left(\begin{bmatrix} A_{1,1} - \lambda & \star & \cdots & \star \\ 0 & A_{2,2} - \lambda & \cdots & \star \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & A_{n,n} - \lambda \end{bmatrix}\right) \\ &= (A_{1,1} - \lambda)(A_{2,2} - \lambda) \cdots (A_{n,n} - \lambda) \end{aligned}$$

has solutions $A_{1,1}, A_{2,2}, \ldots, A_{n,n}$, so the eigenvalues of A are the entries on the main diagonal of A. Hence $\prod_{i=1}^{n} A_{i,i} = \prod_{i=1}^{n} \lambda_i$ completing the proof for upper triangular matrices.

If A is any matrix, the conclusion follows from two facts.

1. The determinant and eigenvalues of $P^{-1}AP$ are the same as the determinant and eigenvalues of A for any invertible $n \times n$ matrix P (see theorem 14).

2. For any $n \times n$ matrix A, there is an $n \times n$ matrix P such that $P^{-1}AP$ is upper triangular (see crumpet 28).

Given these two facts, if $U = P^{-1}AP$, then $\det A = \det U$ and the eigenvalues of A are the eigenvalues of U by fact 1 (theorem 14). Now if P is that special matrix such that U is upper triangular, as guaranteed to exist by fact 2, then the determinant of U (which equals the determinant of A) and the product of the eigenvalues of U (which equals the product of the eigenvalues of A) are both $\prod_{i=1}^{n} U_{i,i}$ and therefore equal. This concludes the proof of the theorem for general matrices.

Crumpet 28: Triangularization

For a square matrix M, $P^{-1}MP$ is a triangularization of M whenever $P^{-1}MP$ is upper triangular. We wish to show that there is a triangularization of any $n \times n$ matrix. Triangularization of a 1×1 matrix is simple enough since all 1×1 matrices are upper triangular. Choose, $P = \begin{bmatrix} 1 \end{bmatrix}$ for example. Proceeding by induction, assume a triangularization exists for every $(k-1) \times (k-1)$ matrix for some $k \geq 2$, and let M be a particular but arbitrary $k \times k$ matrix. Take any eigenpair $\lambda, \mathbf{v}$ of M and find vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-1}$ such that $\{\mathbf{v}, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-1}\}$ is linearly independent. This set can always be found since $\mathbf{v}$ must have at least one nonzero entry ($\mathbf{0}$ is not a permissible eigenvector). Assuming $\mathbf{v}_{i,1} \neq 0$, we may take $\{\mathbf{v}, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-1}\} = \{\mathbf{v}\} \cup \{I_{:,j} : j \neq i\}$. Setting $Q = \begin{bmatrix} \mathbf{v} & \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{n-1} \end{bmatrix}$, Q is invertible (its columns are linearly independent), and

$$\begin{aligned} Q^{-1}MQ &= Q^{-1}M\begin{bmatrix} \mathbf{v} & \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{n-1} \end{bmatrix} \\ &= \begin{bmatrix} Q^{-1}M\mathbf{v} & Q^{-1}M\mathbf{u}_1 & Q^{-1}M\mathbf{u}_2 & \cdots & Q^{-1}M\mathbf{u}_{n-1} \end{bmatrix} \\ &= \begin{bmatrix} \lambda Q^{-1}\mathbf{v} & Q^{-1}M\mathbf{u}_1 & Q^{-1}M\mathbf{u}_2 & \cdots & Q^{-1}M\mathbf{u}_{n-1} \end{bmatrix}. \end{aligned}$$

While we cannot say much about $Q^{-1}M\mathbf{u}_j$ for any j, we can say $\lambda Q^{-1}\mathbf{v} = \lambda I_{:,1}$ because $Q^{-1}Q = Q^{-1}\begin{bmatrix} \mathbf{v} & \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{n-1} \end{bmatrix} = I$. Q^{-1} times the first column of Q must be the first column of I. Hence we have

$$Q^{-1}MQ = \begin{bmatrix} \lambda & \star & \star & \star & \star \\ 0 & \star & \star & \star & \star \\ 0 & \star & \star & \star & \star \\ \vdots & \star & \star & \star & \star \\ 0 & \star & \star & \star & \star \end{bmatrix}.$$

By the inductive hypothesis, there is a triangularization of $(Q^{-1}MQ)_{\backslash 1,1}$. Let R be such that $R^{-1}(Q^{-1}MQ)_{\backslash 1,1}R$ is upper triangular, and set $\hat{Q} = \begin{bmatrix} 1 & 0 \\ 0 & R \end{bmatrix}$. Then $\hat{Q}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & R^{-1} \end{bmatrix}$ and

$$\hat{Q}^{-1}(Q^{-1}MQ)\hat{Q} = \begin{bmatrix} 1 & 0 \\ 0 & R^{-1} \end{bmatrix}\begin{bmatrix} \lambda & \star \\ 0 & (Q^{-1}MQ)_{\backslash 1,1} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & R \end{bmatrix}$$

is upper triangular. Hence $(Q\hat{Q})^{-1}M(Q\hat{Q})$ is a triangularization of M and we set $P = Q\hat{Q}$. This result suffices for our purposes, but the result can be strengthened to specify that $Q\hat{Q}$ have a certain property, a so-called Schur decomposition.

Hence we have two ways to measure the effect of a linear transformation on the plane. In rough terms, a linear transformation expands or compresses areas by a factor equal to the absolute value of the determinant, which is equal to the absolute value of the product of the eigenvalues, of its standard matrix. More precisely a linear transformation expands or compresses areas in the direction of each eigenvector by a factor equal to the absolute value of the associated eigenvalue.

Determinants, eigenvalues, and volumes

The analysis of elementary 3×3 matrices follows much along the same lines as the analysis of 2×2 matrices in section 4.4. Vectors in $\mathbb{R}^3$ can be imagined as arrows or points just as they are in $\mathbb{R}^2$. Images of cubes in $\mathbb{R}^3$ under transformations associated with elementary matrices analogous to the images of the coffee cup in $\mathbb{R}^2$ can be derived. They will also be a collection of reflections, shears, and scalings. Rotation in $\mathbb{R}^3$ can be accomplished by a compostion of scalings and shears just as in $\mathbb{R}^2$. Noninvertible 3×3 matrices can be described by compositions of elementary matrices and projections as well. Hence theorem 11 can be proved for linear operators on $\mathbb{R}^3$. Generally, if the 2's of the present section are replaced by 3's and the word area is replaced by the word volume, the discourse still applies with only minor additional modification. To illustrate, for 3×3 matrices M with eigenvalues $\lambda_1, \lambda_2, \lambda_3$, and three-dimensional regions of space, R,

$$\begin{aligned} \text{volume}(T_M(R)) &= |\det M| \cdot \text{volume}(R) \\ &= |\lambda_1 \lambda_2 \lambda_3| \cdot \text{volume}(R) \end{aligned}$$

and the concluding paragraph in the discussion of transformations of the plane might be rephrased for transformations of space as follows.

> *We have two ways to measure the effect of a linear transformation on space, $\mathbb{R}^3$. In rough terms, a linear transformation expands or compresses volumes by a factor equal to the absolute value of the determinant ,which is equal to the absolute value of the product of the eigenvalues, of its standard matrix. More precisely a linear transformation expands or compresses volumes in the direction of each eigenvector by a factor equal to the absolute value of the associated eigenvalue.*

Crumpet 29: Hyperspace

The main results of this section and the previous are stated and hold for $\mathbb{R}^n$, giving an enterprising individual a basis to extend the ideas of area and volume to dimensions higher than 3! The notion of a hypercube (in hyperspace) is exactly this enterprise.

Affine Transformations

Translations, transformations of the form $T : \mathbb{R}^n \to \mathbb{R}^n$,

$$T(\mathbf{x}) = \mathbf{x} + \mathbf{c},$$

are not linear for any $\mathbf{c} \neq \mathbf{0}$. Can you provide a justification? Answer on page 6.3. But because their geometric effect is to simply displace all points by the same distance and direction, they do not change the shapes of figures and therefore do not change areas or volumes of figures. For a translation T, area$(T(S))$ = area(S) for any set S with measurable area.

Affine transformations, compositions of linear transformations with translations, are consequently not linear either, but their effect on areas is predictable. They scale areas in exactly the same manner as their linear parts. For an affine transformation $F : \mathbb{R}^n \to \mathbb{R}^n$,

$$F(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$$

for some matrix A and vector $\mathbf{c}$ and area$(F(S)) = |\det A| \cdot$ area(S) for any set S with measurable area.

Key Concepts

set image For any transformation (map or function) $f : A \to B$ and subset S of A,

$$f(S) = \{f(s) : s \in S\}$$

determinant and area For any linear transformation $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ and any subset S of $\mathbb{R}^2$ with measurable area,

$$\text{area}(T_A(S)) = |\det A| \cdot \text{area}(S)$$

determinant and volume For any linear transformation $T_A : \mathbb{R}^3 \to \mathbb{R}^3$ and any subset S of $\mathbb{R}^3$ with measurable volume, r

$$\text{volume}(T_A(S)) = |\det A| \cdot \text{volume}(S)$$

determinant and eigenvalues The determinant of any square matrix is the product of its eigenvalues.

triangularization For any square matrix M there is an invertible matrix P such that $P^{-1}MP$ is upper triangular.

affine transformation The composition of a linear transformation with a translation.

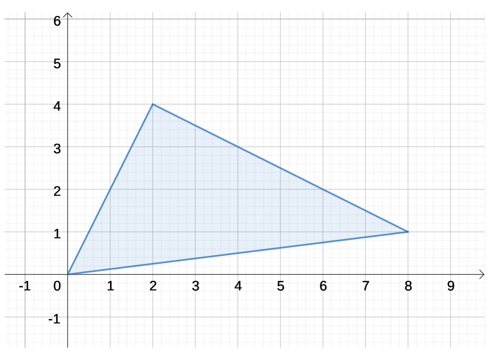
Exercises

1. Find the area of the parallelogram with vertices
   (a) $(0,0)$, $(2,3)$, $(5,-1)$, $(3,-4)$ [𝕊]-328
   (b) $(0,0)$, $(1,8)$, $(-1,5)$, $(-2,-3)$
   (c) $(0,0)$, $(-5,6)$, $(7,18)$, $(12,12)$
   (d) $(4,5)$, $(8,11)$, $(16,12)$, $(12,6)$ [𝕊]-329
   (e) $(-1,3)$, $(3,-1)$, $(9,-4)$, $(5,0)$
   (f) $(4,-2)$, $(11,-5)$, $(9,-10)$, $(2,-7)$

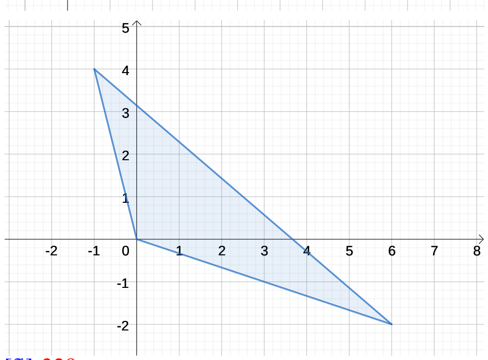
2. Use the fact that $\text{area}(T_A(S)) = |\det A| \cdot \text{area}(S)$ to justify the claim that the area of the parallelogram determined by the columns of a 2×2 matrix A is $|\det A|$. Alternatively, justify the claim that the area of the parallelogram determined by two vectors (anchored at the origin) is the absolute value of the determinant of the matrix whose columns are the two vectors.

3. Calculate the area of the triangle as half of a determinant. See exercise 2 for a hint.

(a)

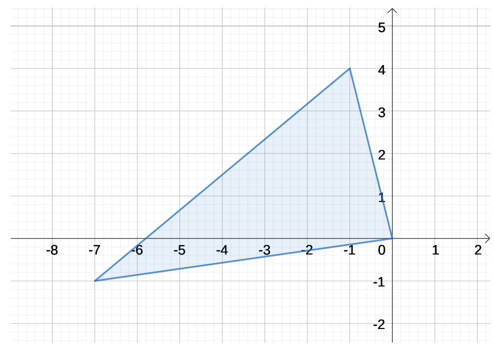


(b)

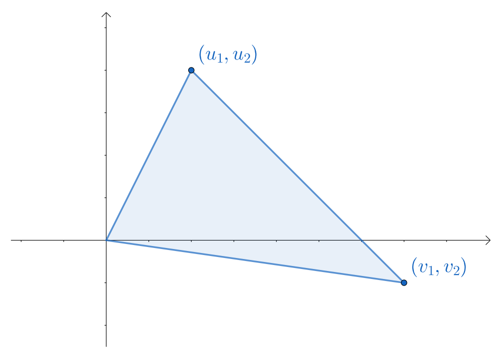


[𝕊]-329

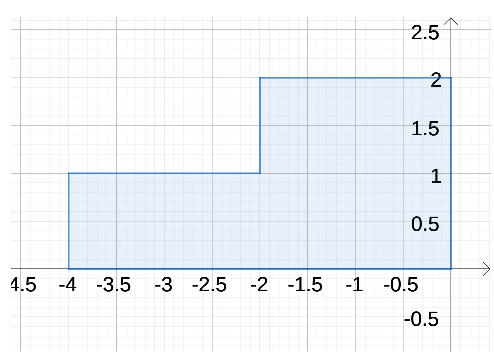
(c)



(d)



4. The image of the hexagon with adjacent vertices $(0,0)$, $(2,0)$, $(2,1)$, $(1,1)$, $(1,2)$, $(0,2)$ under the transformation $T(\mathbf{x}) = A\mathbf{x}$ where A is a 2×2 matrix is shown. What is the absolute value of the determinant of A?
   [𝔸]-361



5. What are the eigenvalues of the matrix A of question 4 assuming no reflection? [𝔸]-361

6. Suppose M factors as $P^{-1}UP$. Use the information to find (i) $\det M$ and (ii) the eigenvalues of M.

   (a) $U = \begin{bmatrix} -4 & -9 \\ 0 & 5 \end{bmatrix}$; $P = \begin{bmatrix} 12 & -10 \\ 10 & 4 \end{bmatrix}$
   [𝕊]-329

   (b) $U = \begin{bmatrix} -6 & 12 \\ 0 & -8 \end{bmatrix}$; $P = \begin{bmatrix} -12 & 10 \\ -7 & 8 \end{bmatrix}$

   (c) $U = \begin{bmatrix} -11 & -4 & -7 \\ 0 & 0 & -12 \\ 0 & 0 & 12 \end{bmatrix}$;
   $P = \begin{bmatrix} 4 & 9 & 11 \\ -5 & 1 & -6 \\ -7 & -3 & 6 \end{bmatrix}$

   (d) $U = \begin{bmatrix} -11 & 4 & 11 \\ 0 & -5 & -2 \\ 0 & 0 & 3 \end{bmatrix}$;
   $P = \begin{bmatrix} 12 & 8 & -12 \\ -6 & -10 & 1 \\ 6 & -11 & 4 \end{bmatrix}$

7. Use SageMath to verify your answers in question 6 by calculating M and having SageMath compute the determinant and eigenvalues.

   (a) SageMathCell 94 [𝕊]-329

   (b) SageMathCell 95

   (c) SageMathCell 96

   (d) SageMathCell 97

8. Use the fact that $\text{volume}(T_A(S)) = |\det A| \cdot \text{volume}(S)$ to justify the claim that the volume of the parallelepiped determined by the columns of a 3×3 matrix A is $\det A$.

9. Let S be the unit square with opposite corners $(0,0)$ and $(1,1)$. Sketch the image of S under the affine transformation $T(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$.

   (a) $A = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}$; $\mathbf{c} = \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}$

   (b) $A = \begin{bmatrix} 0 & -1/2 \\ 1/2 & 0 \end{bmatrix}$; $\mathbf{c} = \begin{bmatrix} 1 \\ 1/2 \end{bmatrix}$ [𝔸]-361

   (c) $A = \begin{bmatrix} 0 & 1/2 \\ -1/2 & 0 \end{bmatrix}$; $\mathbf{c} = \begin{bmatrix} 0 \\ 1/2 \end{bmatrix}$

10. Let S be the triangle with vertices $(0,0)$, $(1,1)$, and $(-1,1)$ and suppose F and G are affine transformations such that $F(S)$ is the triangle with vertices $(0,0)$, $(0,1)$, and $(-1,1)$; and $G(S)$ is the triangle with vertices $(0,0)$, $(1,1)$, and $(0,1)$. Draw S, $F(S)$, and $G(S)$ and use your sketch to determine the determinants of the linear parts of F and G?

11. Crumpet 28 outlines a recursive procedure for triangularizing any matrix. It is constructive, giving an algorithm for finding the traignularizing matrix P. Use it to find P such that $P^{-1}MP$ is upper triangular. One eigenvalue of M is given.

    (a) $M = \begin{bmatrix} 13 & -8 \\ 20 & -13 \end{bmatrix}$; $\lambda = -3$ [𝕊]-330

    (b) $M = \begin{bmatrix} 22 & -7 \\ 28 & -13 \end{bmatrix}$; $\lambda = -6$

    (c) $M = \begin{bmatrix} 11 & -8 \\ 2 & 1 \end{bmatrix}$; $\lambda = 3$ [𝔸]-361

    (d) $M = \begin{bmatrix} -9 & 7 & 28 \\ -11 & -27 & -28 \\ 8 & 8 & 12 \end{bmatrix}$; $\lambda = 12$

    (e) $M = \begin{bmatrix} -34 & -50 & -24 \\ 37 & 53 & 24 \\ -26 & -35 & -14 \end{bmatrix}$; $\lambda = 4$ [𝔸]-361

    (f) $M = \begin{bmatrix} 78 & -50 & -54 \\ -167 & 135 & 186 \\ 201 & -150 & -193 \end{bmatrix}$; $\lambda = 5$
    [𝔸]-330

12. Prove that any square matrix M can be factored as PUP^{-1} for some invertible matrix P and upper triangular matrix U.

13. Suppose M is invertible. What can you say about the eigenvalues of $M^T M$? HINT: see exercise 2 of section 3.7.

Answers

further analysis The parallelogram determined by $\mathbf{v}_1$ and $\mathbf{v}_2$ is the set $S = \{\beta\mathbf{v}_1 + \alpha\mathbf{v}_2 : 0 \le \alpha, \beta \le 1\}$ so its image is

$$\begin{aligned} T_A(S) &= T_A\left(\{\beta\mathbf{v}_1 + \alpha\mathbf{v}_2 : 0 \le \alpha,\beta \le 1\}\right) \\ &= \{T_A(\beta\mathbf{v}_1 + \alpha\mathbf{v}_2) : 0 \le \alpha,\beta \le 1\} \\ &= \{\beta T_A(\mathbf{v}_1) + \alpha T_A(\mathbf{v}_2) : 0 \le \alpha,\beta \le 1\} \\ &= \{\beta\lambda_1\mathbf{v}_1 + \alpha\lambda_2\mathbf{v}_2 : 0 \le \alpha,\beta \le 1\} \end{aligned}$$

which is the paralelogram determined by $T_A(\mathbf{v}_1)$ and $T_A(\mathbf{v}_2)$.

translations are not linear On the one hand,

$$T(\mathbf{x} + \mathbf{y}) = \mathbf{x} + \mathbf{y} + \mathbf{c}$$

and on the other hand,

$$\begin{aligned} T(\mathbf{x}) + T(\mathbf{y}) &= (\mathbf{x} + \mathbf{c}) + (\mathbf{y} + \mathbf{c}) \\ &= \mathbf{x} + \mathbf{y} + 2\mathbf{c} \end{aligned}$$

so $T(\mathbf{x} + \mathbf{y}) \neq T(\mathbf{x}) + T(\mathbf{y})$ whenever $\mathbf{c} \neq \mathbf{0}$.

6.4 Approximation [4.1, 4.6, 5.1, 5.3]

From the very beginning of our discussion of linear systems, we acknowledged that there were systems with no solution (see section 2.1 exercise 4). This was a familiar state of affairs as you undoubtedly have seen equations like $x^2 + 1 = 0$, $\sin\theta = 2$, and $\frac{1}{2+e^t} = 3$, all of which have "no solution". Full disclosure, if your instructor or textbook claimed equations such as these had no solution, what they meant was *no real number solution*. All three equations have complex number solutions. $\sqrt{-1}$, $\sin^{-1}(2)$, and $\ln\left(-\frac{5}{3}\right)$ are perfectly well defined complex numbers, and are, respectively, solutions of the three equations. It's possible you studied complex numbers enough to know this already, but it's also possible this comes as a revelation. No worries either way.

Linear systems with no solution are different. When we say they have no solution, they have no integer solution, no rational number solution, no real number solution, and no complex number solution. They simply have no solution. What more is there to say?

The linear equation

$$54x + 30y = 17 \tag{6.4.1}$$

has *no integer solution*. This can be seen by factoring a 6 from the left-hand side:

$$6(9x + 5y) = 17$$

showing that the left side is, for any integers x and y, a multiple of 6 while the right side is not. The best we can hope for are integers x and y that make $54x + 30y$ close to 17. To say it another way, we can look for integers x and y so that

$$|(54x + 30y) - 17|$$

(the distance between $54x + 30y$ and 17) is small. In fact, if we could find a minimum of this quantity, that would mean something. Among all the pairs of integers x and y, this pair (or these pairs) make $54x + 30y$ as close to 17 as possible. Can you find the minimum possible value of $|(54x + 30y) - 17|$ for integers x and y? Answer on page 224.

$54(-34) + 30(78) = 18$, so $(\hat{x}, \hat{y}) = (-34, 78)$ is a **best approximation** of an integer solution of (6.4.1). The pair $(-34, 78)$ does not solve (6.4.1), but it makes its two sides as close as possible using integers. That is,

$$|(54\hat{x} + 30\hat{y}) - 17| \le |(54x + 30y) - 17|$$

for all integer pairs (x, y). Even when an equation has no solution, it may have a best approximation.

Using this discussion as a model for linear systems with no solution, we ask whether inconsistent systems

$$M\mathbf{v} = \mathbf{b}$$

have a best approximation. That is, can we find $\hat{\mathbf{v}}$ such that

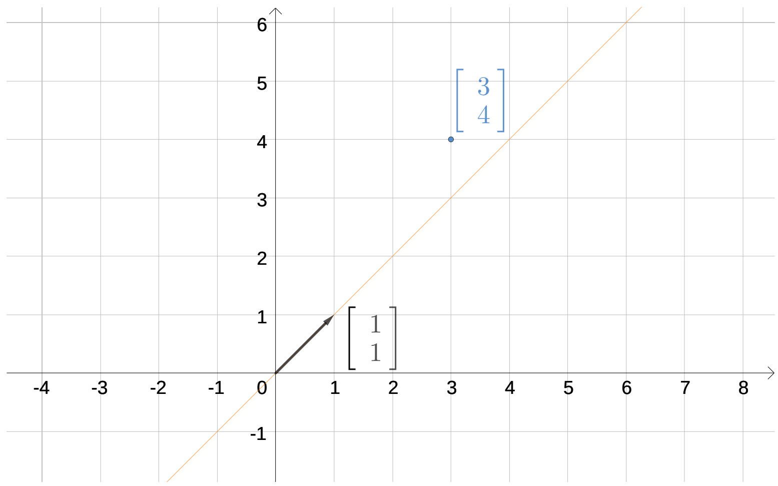
$$\|M\hat{\mathbf{v}} - \mathbf{b}\| \le \|M\mathbf{v} - \mathbf{b}\|$$

for all $\mathbf{v}$? For example,

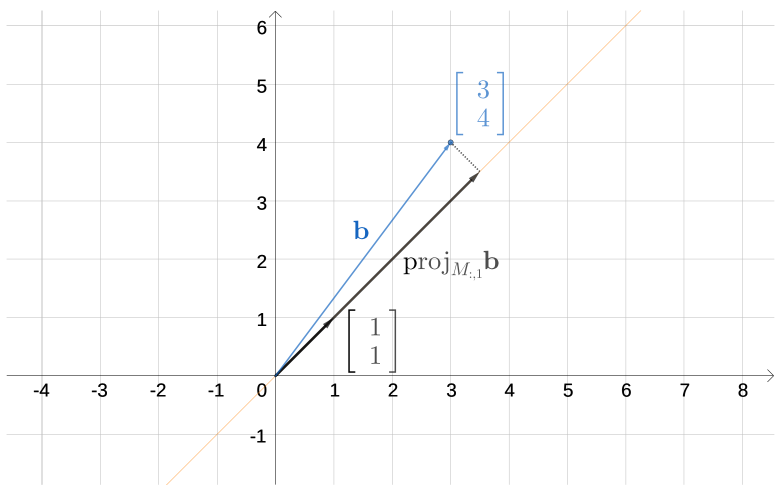
$$\begin{bmatrix} 1 & -3 \\ 4 & 9 \\ -9 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \\ 7 \end{bmatrix} \tag{6.4.2}$$

is inconsistent. Can you show this? Answer on page 224. To say it another way, $\begin{bmatrix} 8 \\ 5 \\ 7 \end{bmatrix}$ is not in the column space of $\begin{bmatrix} 1 & -3 \\ 4 & 9 \\ -9 & 6 \end{bmatrix}$, hinting at how to find a best approximation—look inside the column space of $\begin{bmatrix} 1 & -3 \\ 4 & 9 \\ -9 & 6 \end{bmatrix}$ for a vector that is as close to $\begin{bmatrix} 8 \\ 5 \\ 7 \end{bmatrix}$ as possible.

Actually, we have done this to some extent already! $\mathbf{b} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ is not in the column space of $M = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ since $\mathbf{b}$ is not a multiple (linear combination) of $M_{:,1}$. Nonetheless there is a multiple (linear combination) of $M_{:,1}$ that is closest to $\mathbf{b}$. Geometrically, this means there is a point on the line determined by $M_{:,1}$ closest to $\mathbf{b}$. This situation is diagrammed here.



We know the shortest distance between a point and a line is measured perpendicularly. The point on the line where this shortest distance occurs coincides exactly with the orthogonal projection of $\mathbf{b}$ onto $M_{:,1}$, as diagrammed here.



Helpful to this discussion is to see orthogonal projection as projecting a vector onto a subspace (rather than a vector). The line determined by $M_{:,1}$ is a subspace of $\mathbb{R}^2$ as it is the span of $M_{:,1}$.

In three dimensions, there is a point in a plane nearest any point/vector not in that plane. That point occurs exactly at the projection of the vector onto the plane (and that plane is a subspace of $\mathbb{R}^3$). Again, projection is best viewed as projecting a vector onto a subspace.

With this in mind, we have to address the questions of (i) how to project a vector onto a subspace of dimension greater than one and (ii) whether that projection is always the nearest point/vector within the subspace. A lot of this work has already been done, but there is a bit more to do now. Question 22 of section 5.3 provides a good backdrop for this conversation.

First, let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ be an orthogonal basis for a subspace W of an inner product space V. Then the **orthogonal projection** of any $\mathbf{v}$ in V onto W, denoted $\mathrm{proj}_W\mathbf{v}$, is defined by

$$\mathrm{proj}_W\mathbf{v} = \mathrm{proj}_{\mathbf{b}_1}\mathbf{v} + \mathrm{proj}_{\mathbf{b}_2}\mathbf{v} + \cdots + \mathrm{proj}_{\mathbf{b}_p}\mathbf{v}.$$

Because each projection is a multiple of one of the basis elements, this is a linear combination of the basis elements and therefore lies in W. Next, we will need some terminology.

If W is a subspace of an inner product space V and $\mathbf{v}$ is orthogonal to every vector in W, then we say $\mathbf{v}$ is **orthogonal** to W. The set of all vectors in V orthogonal to W is called the **orthogonal complement** of W and is denoted $W^\perp$ (read "W perp"). Can you show that $W^\perp$ is a subspace of V? Answer on page 224.

Just as $\mathbf{v}$ and $\mathbf{v} - \mathrm{proj}_{\mathbf{w}}\mathbf{v}$ are orthogonal for any vectors $\mathbf{v}$ and $\mathbf{w}$ of an inner product space V (see section 5.3), we can now show that $\mathbf{v}$ and $\mathbf{v} - \mathrm{proj}_W\mathbf{v}$ are orthogonal for any vector $\mathbf{v}$ and subspace W of an inner product space V.

Letting $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ be an *orthogonal basis* of W,

$$\begin{aligned}
\langle \mathbf{v} - \mathrm{proj}_W \mathbf{v}, \mathbf{b}_j \rangle &= \left\langle \mathbf{v} - \left(\mathrm{proj}_{\mathbf{b}_1} \mathbf{v} + \mathrm{proj}_{\mathbf{b}_p} \mathbf{v} + \cdots + \mathrm{proj}_{\mathbf{b}_p} \mathbf{v} \right), \mathbf{b}_j \right\rangle \\
&= \left\langle \mathbf{v}, \mathbf{b}_j \right\rangle - \left\langle \mathrm{proj}_{\mathbf{b}_1} \mathbf{v}, \mathbf{b}_j \right\rangle - \left\langle \mathrm{proj}_{\mathbf{b}_p} \mathbf{v}, \mathbf{b}_j \right\rangle - \cdots - \left\langle \mathrm{proj}_{\mathbf{b}_p} \mathbf{v}, \mathbf{b}_j \right\rangle \\
&= \left\langle \mathbf{v}, \mathbf{b}_j \right\rangle - \left\langle \mathrm{proj}_{\mathbf{b}_j} \mathbf{v}, \mathbf{b}_j \right\rangle \\
&= \left\langle \mathbf{v}, \mathbf{b}_j \right\rangle - \left\langle \frac{\langle \mathbf{v}, \mathbf{b}_j \rangle}{\langle \mathbf{b}_j, \mathbf{b}_j \rangle} \mathbf{b}_j, \mathbf{b}_j \right\rangle \\
&= \left\langle \mathbf{v}, \mathbf{b}_j \right\rangle - \frac{\langle \mathbf{v}, \mathbf{b}_j \rangle}{\langle \mathbf{b}_j, \mathbf{b}_j \rangle} \left\langle \mathbf{b}_j, \mathbf{b}_j \right\rangle \\
&= 0
\end{aligned}$$

for each $j = 1, 2, \ldots, p$, so $\mathbf{v} - \mathrm{proj}_W \mathbf{v}$ is orthogonal to every element of a basis of W. This is enough to show that $\mathbf{v} - \mathrm{proj}_W \mathbf{v}$ is orthogonal to every vector in W and therefore $\mathbf{v} - \mathrm{proj}_W \mathbf{v}$ is in $W^\perp$. Can you provide the details? Answer on page 224. This leads to the following theorem.

Theorem 16. [Orthogonal Decomposition] *Let W be a subspace of an inner product space V. Each $\mathbf{v} \in V$ can be written uniquely as a sum*

$$\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$$

where $\mathbf{w} \in W$ and $\mathbf{w}^\perp \in W^\perp$.

Existence: we have just shown that $\mathbf{v} - \mathrm{proj}_W \mathbf{v}$ is in $W^\perp$. Since $\mathrm{proj}_W \mathbf{v}$ is in W and

$$\mathbf{v} = \mathrm{proj}_W \mathbf{v} + (\mathbf{v} - \mathrm{proj}_W \mathbf{v})$$

we have existence.
Uniqueness: suppose $\mathbf{v} = \hat{\mathbf{w}} + \hat{\mathbf{w}}^\perp$ for some (possibly other) $\hat{\mathbf{w}}$ in W and $\hat{\mathbf{w}}^\perp$ in $W^\perp$. Then, of course, $\mathbf{w} + \mathbf{w}^\perp = \hat{\mathbf{w}} + \hat{\mathbf{w}}^\perp$ so $\mathbf{w} - \hat{\mathbf{w}} = \hat{\mathbf{w}}^\perp - \mathbf{w}^\perp$. Noting that $\mathbf{w} - \hat{\mathbf{w}}$ is in W and $\hat{\mathbf{w}}^\perp - \mathbf{w}^\perp$ is in $W^\perp$, we have $\langle \mathbf{w} - \hat{\mathbf{w}}, \hat{\mathbf{w}}^\perp - \mathbf{w}^\perp \rangle = 0$. Setting $\mathbf{x} = \mathbf{w} - \hat{\mathbf{w}} = \hat{\mathbf{w}}^\perp - \mathbf{w}^\perp$, this means

$$\langle \mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{w} - \hat{\mathbf{w}}, \hat{\mathbf{w}}^\perp - \mathbf{w}^\perp \rangle = 0,$$

so $\mathbf{x} = \mathbf{0}$ and therefore $\mathbf{w} = \hat{\mathbf{w}}$ and $\hat{\mathbf{w}}^\perp = \mathbf{w}^\perp$.

Corollary 17. $\mathbf{v} = \mathit{proj}_W \mathbf{v} + (\mathbf{v} - \mathit{proj}_W \mathbf{v})$ *is the unique decomposition of* $\mathbf{v}$ *into the sum of two vectors, one in W and one in $W^\perp$.*

Finally, we are ready to answer our original question. In the form of a theorem, we have the following.

Theorem 18. [Best Approximation] *If W is a subspace of an inner product space V and $\mathbf{v}$ is in V, then $\mathbf{w} = \mathit{proj}_W \mathbf{v}$ is the closest vector to $\mathbf{v}$ in W.*

Justification of theorem 18 relies on a generalization of the Pythagorean theorem. Can you prove that if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal (vectors of an inner product space), then $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$? Answer on page 225. Now let $\hat{\mathbf{w}}$ be any vector in W, $\hat{\mathbf{w}} \neq \mathbf{w}$. Since $\mathbf{v} - \mathbf{w}$ is in $W^\perp$ and $\mathbf{w} - \hat{\mathbf{w}}$ is in W, they are orthogonal, and the Pythagorean applies. But $(\mathbf{v} - \mathbf{w}) + (\mathbf{w} - \hat{\mathbf{w}}) = \mathbf{v} - \hat{\mathbf{w}}$ so

$$\|\mathbf{v} - \hat{\mathbf{w}}\|^2 = \|\mathbf{v} - \mathbf{w}\|^2 + \|\mathbf{w} - \hat{\mathbf{w}}\|^2 .$$

Since $\hat{\mathbf{w}} \neq \mathbf{w}$, $\|\mathbf{w} - \hat{\mathbf{w}}\|^2 > 0$ and therfore $\|\mathbf{v} - \hat{\mathbf{w}}\|^2 > \|\mathbf{v} - \mathbf{w}\|^2$. In other words, $\mathbf{w}$ is the closest point to $\mathbf{v}$ in W.

Corollary 19. [Best Approximation for a Linear System] *Given any $m \times n$ matrix M and vector $\mathbf{b}$ in $\mathbb{R}^m$, let W be the column space of M. Then any solution of $M\hat{\mathbf{v}} = \mathit{proj}_W \mathbf{b}$ for $\hat{\mathbf{v}}$ is a best approximation to a solution of $M\mathbf{v} = \mathbf{b}$.*

Can you use theorem 18 to prove theorem 19? Answer on page 225. Finally, we can return to (6.4.2) and provide an answer. We need to project

$$\mathbf{b} = \begin{bmatrix} 8 \\ 5 \\ 7 \end{bmatrix}$$

onto W, the column space of

$$M = \begin{bmatrix} 1 & -3 \\ 4 & 9 \\ -9 & 6 \end{bmatrix}.$$

This requires an orthogonal basis of the column space of M. Using Gram-Schmidt orthogonalization, let

$$\mathbf{w}_1 = M_{:,2} = \begin{bmatrix} -3 \\ 9 \\ 6 \end{bmatrix}$$

and

$$\begin{aligned} \mathbf{w}_2 = M_{:,1} - \mathrm{proj}_{\mathbf{w}_1} M_{:,1} &= \begin{bmatrix} 1 \\ 4 \\ -9 \end{bmatrix} - \frac{\langle M_{:,1}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle} \begin{bmatrix} -3 \\ 9 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ -9 \end{bmatrix} - \frac{-1}{6} \begin{bmatrix} -3 \\ 9 \\ 6 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 1 \\ 11 \\ -16 \end{bmatrix}. \end{aligned}$$

Taking (scalar multiples of $\mathbf{w}_1$ and $\mathbf{w}_2$)

$$\{\mathbf{b}_1, \mathbf{b}_2\} = \left\{ \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 11 \\ -16 \end{bmatrix} \right\}$$

as the orthogonal basis of the column space of M, the projection of $\mathbf{b}$ onto W is

$$\begin{aligned} \mathrm{proj}_W \mathbf{b} = \mathrm{proj}_{\mathbf{b}_1} \mathbf{b} + \mathrm{proj}_{\mathbf{b}_2} \mathbf{b} &= \frac{\langle \mathbf{b}, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle} \mathbf{b}_1 + \frac{\langle \mathbf{b}, \mathbf{b}_2 \rangle}{\langle \mathbf{b}_2, \mathbf{b}_2 \rangle} \mathbf{b}_2 \\ &= \frac{3}{2} \mathbf{b}_1 - \frac{7}{54} \mathbf{b}_2 = \frac{1}{27} \begin{bmatrix} -44 \\ 83 \\ 137 \end{bmatrix}. \end{aligned}$$

Hence

$$\frac{1}{27} \begin{bmatrix} -44 \\ 83 \\ 137 \end{bmatrix} \approx \begin{bmatrix} -1.63 \\ 3.07 \\ 5.07 \end{bmatrix}$$

is the closest vector to

$$\begin{bmatrix} 8 \\ 5 \\ 7 \end{bmatrix}$$

in the column space of M. The distance between these two vectors happens to be

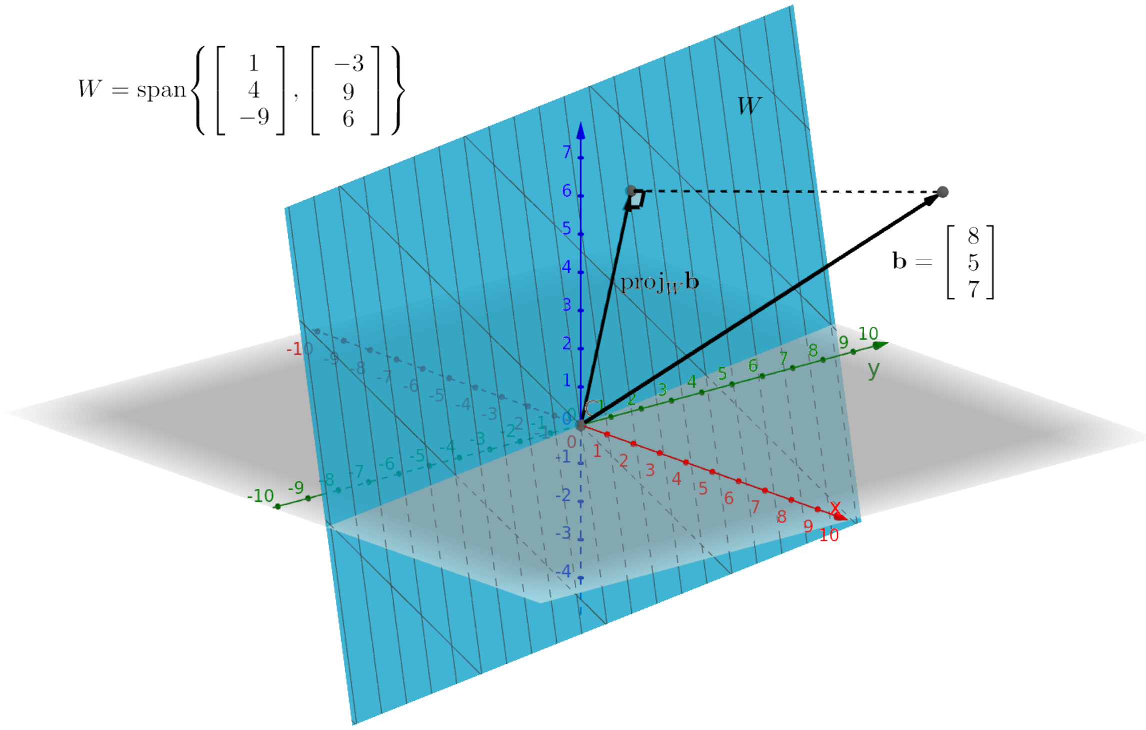
$$\begin{aligned} \left\| \frac{1}{27} \begin{bmatrix} -44 \\ 83 \\ 137 \end{bmatrix} - \begin{bmatrix} 8 \\ 5 \\ 7 \end{bmatrix} \right\| &= \frac{1}{27} \left\| \begin{bmatrix} -44 \\ 83 \\ 137 \end{bmatrix} - \begin{bmatrix} 8 \\ 5 \\ 7 \end{bmatrix} \right\| = \frac{1}{27} \left\| \begin{bmatrix} -52 \\ 78 \\ 130 \end{bmatrix} \right\| \\ &= \frac{26}{27} \sqrt{38} \approx 5.936 \end{aligned}$$

and there is no vector in the column space of M closer to $\mathbf{b}$. Hence the solution of $M\hat{\mathbf{v}} = \mathrm{proj}_W \mathbf{b}$, $\begin{bmatrix} 1 & -3 \\ 4 & 9 \\ -9 & 6 \end{bmatrix} \hat{\mathbf{v}} = \frac{1}{27} \begin{bmatrix} -44 \\ 83 \\ 137 \end{bmatrix}$, gives the best approximation of a solution:

$$\begin{bmatrix} 1 & -3 & -\frac{44}{27} \\ 4 & 9 & \frac{83}{27} \\ -9 & 6 & \frac{137}{27} \end{bmatrix} \rightarrow$$

See figure 6.4.1.

Figure 6.4.1: Best Approximation of a Solution of $\begin{bmatrix} 1 & -3 \\ 4 & 9 \\ -9 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \\ 7 \end{bmatrix}$



Key Concepts

best approximation of $M\mathbf{v} = \mathbf{b}$ is a vector $\hat{\mathbf{v}}$ such that

$$\|M\hat{\mathbf{v}} - \mathbf{b}\| < \|M\mathbf{v} - \mathbf{b}\|$$

for all $\mathbf{v} \neq \hat{\mathbf{v}}$. ($M$ is an $m \times n$ matrix, $\mathbf{v}, \hat{\mathbf{v}}$ are in $\mathbb{R}^n$ and $\mathbf{b}$ is in $\mathbb{R}^m$.)

best approximation theorem see theorem 18 and corollary 19.

orthogonal a vector $\mathbf{v}$ is orthogonal to a subspace W if $\mathbf{v}$ is orthogonal to every vector in W. $\mathbf{v}$ is orthogonal to W if and only if $\mathbf{v}$ is orthogonal to every vector in a basis of W.

orthogonal projection of a vector $\mathbf{v}$ onto a subspace W, denoted $\operatorname{proj}_W \mathbf{v}$, is defined by

$$\operatorname{proj}_W \mathbf{v} = \operatorname{proj}_{\mathbf{b}_1} \mathbf{v} + \operatorname{proj}_{\mathbf{b}_2} \mathbf{v} + \cdots + \operatorname{proj}_{\mathbf{b}_p} \mathbf{v}$$

for any orthogonal basis $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ of W.

orthogonal complement of a subspace W is the set of all vectors orthogonal to W, denoted $W^\perp$. For any vector $\mathbf{v}$, $\mathbf{v} - \operatorname{proj}_W \mathbf{v}$ is in $W^\perp$.

orthogonal decomposition writing $\mathbf{v}$ as a sum $\mathbf{w} + \mathbf{w}^\perp$ where $\mathbf{w}$ is in W and $\mathbf{w}^\perp$ is in $W^\perp$. See theorem 16.

Exercises

1. What multiple of $\mathbf{v}$ lands closest to the point (is a best approximation)?

   (a) $\mathbf{v} = \begin{bmatrix} 10 \\ 11 \end{bmatrix}$; $(5, -5)$

   (b) $\mathbf{v} = \begin{bmatrix} 7 \\ -6 \end{bmatrix}$; $(12, 1)$ [$\mathbb{S}$]-331

(c) $\mathbf{v} = \begin{bmatrix} 10 \\ 6 \end{bmatrix}$; $(1,5)$

(d) $\mathbf{v} = \begin{bmatrix} 7 \\ 1 \\ 5 \end{bmatrix}$; $(0,-7,12)$ [𝔸]-361

(e) $\mathbf{v} = \begin{bmatrix} -5 \\ 5 \\ -7 \end{bmatrix}$; $(0,-1,3)$

(f) $\mathbf{v} = \begin{bmatrix} 12 \\ 2 \\ -11 \end{bmatrix}$; $(-4,9,-12)$ [𝔸]-361

(g) $\mathbf{v} = \begin{bmatrix} 2 \\ -11 \\ 0 \\ -3 \end{bmatrix}$; $(0,-9,7,-4)$

(h) $\mathbf{v} = \begin{bmatrix} -1 \\ -7 \\ 11 \\ 12 \end{bmatrix}$; $(-10,6,3,-4)$ [𝔸]-361

(i) $\mathbf{v} = \begin{bmatrix} 8 \\ -5 \\ -6 \\ -9 \end{bmatrix}$; $(-1,12,-8,-2)$

2. Use orthogonal projection to find the distance between the point and the line ℓ.

   (a) $(-10,12)$; $\ell(x) = -\frac{11}{10}x$

   (b) $(-3,-4)$; $\ell(x) = \frac{1}{5}x$ [𝕊]-331

   (c) $(-8,9)$; $\ell(x) = 8x$

   (d) $(8,11)$; $\ell = \left\{ r\begin{bmatrix} -3 \\ 0 \end{bmatrix} : r \text{ in } \mathbb{R} \right\}$ [𝔸]-361

   (e) $(5,-10)$; $\ell = \left\{ r\begin{bmatrix} -1 \\ 1 \end{bmatrix} : r \text{ in } \mathbb{R} \right\}$

   (f) $(6,-7,5)$; $\ell = \left\{ r\begin{bmatrix} -3 \\ -11 \\ 0 \end{bmatrix} : r \text{ in } \mathbb{R} \right\}$
   [𝔸]-361

   (g) $(-4,1,12)$; $\ell = \left\{ r\begin{bmatrix} 5 \\ 0 \\ 9 \end{bmatrix} : r \text{ in } \mathbb{R} \right\}$

   (h) $(-2,0,2,-1)$; $\ell = \left\{ r\begin{bmatrix} -4 \\ -11 \\ -8 \\ 0 \end{bmatrix} : r \text{ in } \mathbb{R} \right\}$
   [𝔸]-361

   (i) $(6,11,5,-8)$; $\ell = \left\{ r\begin{bmatrix} 5 \\ -12 \\ -8 \\ 2 \end{bmatrix} : r \text{ in } \mathbb{R} \right\}$

3. Find the orthogonal projection of $\mathbf{v}$ onto span$\mathcal{B}$. $\mathcal{B}$ is an orthogonal set.

   (a) $\mathbf{v} = \begin{bmatrix} -8 \\ -9 \end{bmatrix}$; $\mathcal{B} = \left\{ \begin{bmatrix} 10 \\ -6 \end{bmatrix} \right\}$ [𝔸]-361

   (b) $\mathbf{v} = \begin{bmatrix} -1 \\ -7 \end{bmatrix}$; $\mathcal{B} = \left\{ \begin{bmatrix} 4 \\ 8 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\}$

   (c) $\mathbf{v} = \begin{bmatrix} -4 \\ -6 \\ 9 \end{bmatrix}$; $\mathcal{B} = \left\{ \begin{bmatrix} 4 \\ 0 \\ -10 \end{bmatrix} \right\}$ [𝔸]-361

   (d) $\mathbf{v} = \begin{bmatrix} 12 \\ 5 \\ -10 \end{bmatrix}$; $\mathcal{B} = \left\{ \begin{bmatrix} -7 \\ 10 \\ 2 \end{bmatrix}, \begin{bmatrix} -6 \\ -4 \\ -1 \end{bmatrix} \right\}$

   (e) $\mathbf{v} = \begin{bmatrix} 9 \\ -12 \\ 5 \end{bmatrix}$; $\mathcal{B} = \left\{ \begin{bmatrix} -1 \\ -4 \\ 5 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$
   [𝕊]-331

   (f) $\mathbf{v} = \begin{bmatrix} 11 \\ -12 \\ 4 \\ 12 \end{bmatrix}$; $\mathcal{B} = \left\{ \begin{bmatrix} 7 \\ 8 \\ -10 \\ 11 \end{bmatrix} \right\}$

   (g) $\mathbf{v} = \begin{bmatrix} 11 \\ 0 \\ -12 \\ 5 \end{bmatrix}$; $\mathcal{B} = \left\{ \begin{bmatrix} 12 \\ 1 \\ -4 \\ 3 \end{bmatrix}, \begin{bmatrix} 6 \\ -11 \\ 10 \\ -7 \end{bmatrix} \right\}$
   [𝔸]-361

   (h) $\mathbf{v} = \begin{bmatrix} -2 \\ -8 \\ 8 \\ 8 \end{bmatrix}$; $\mathcal{B} = \left\{ \begin{bmatrix} -2 \\ -6 \\ -2 \\ 4 \end{bmatrix}, \begin{bmatrix} 12 \\ -1 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 26 \\ -7 \\ 81 \\ 43 \end{bmatrix} \right\}$

4. Find the orthogonal projection of $\mathbf{v}$ onto span$S$. S is not an orthogonal set.

   (a) $\mathbf{v} = \begin{bmatrix} 5 \\ -9 \end{bmatrix}$; $S = \left\{ \begin{bmatrix} 4 \\ -1 \end{bmatrix}, \begin{bmatrix} -12 \\ 3 \end{bmatrix} \right\}$

   (b) $\mathbf{v} = \begin{bmatrix} 7 \\ -7 \\ 9 \end{bmatrix}$;
   $S = \left\{ \begin{bmatrix} 0 \\ -9 \\ -6 \end{bmatrix}, \begin{bmatrix} 6 \\ -16 \\ 2 \end{bmatrix}, \begin{bmatrix} 6 \\ -7 \\ 8 \end{bmatrix} \right\}$ [𝕊]-332

   (c) $\mathbf{v} = \begin{bmatrix} 8 \\ 9 \\ -9 \end{bmatrix}$;
   $S = \left\{ \begin{bmatrix} -8 \\ -5 \\ 12 \end{bmatrix}, \begin{bmatrix} -11 \\ -10 \\ -9 \end{bmatrix}, \begin{bmatrix} 7 \\ 5 \\ 11 \end{bmatrix} \right\}$

   (d) $\mathbf{v} = \begin{bmatrix} -8 \\ -1 \\ 8 \\ 5 \end{bmatrix}$; $S = \left\{ \begin{bmatrix} 9 \\ -5 \\ -10 \\ 10 \end{bmatrix}, \begin{bmatrix} -12 \\ -2 \\ 12 \\ -9 \end{bmatrix} \right\}$
   [𝔸]-361

   (e) $\mathbf{v} = \begin{bmatrix} 4 \\ 5 \\ 3 \\ -8 \end{bmatrix}$;
   $S = \left\{ \begin{bmatrix} -5 \\ 1 \\ 11 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ 12 \\ 7 \\ 11 \end{bmatrix}, \begin{bmatrix} -3 \\ 9 \\ -6 \\ 7 \end{bmatrix} \right\}$

(f) $\mathbf{v} = \begin{bmatrix} 6 \\ 10 \\ 0 \\ 1 \end{bmatrix}$;
$S = \left\{ \begin{bmatrix} -5 \\ 9 \\ -10 \\ -11 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 5 \\ 8 \end{bmatrix}, \begin{bmatrix} -9 \\ -4 \\ -12 \\ -7 \end{bmatrix}, \begin{bmatrix} 10 \\ -2 \\ -12 \\ -11 \end{bmatrix} \right\}$
[𝔸]-361

5. Is $\mathbf{v}$ in $W^\perp$?

(a) $\mathbf{v} = \begin{bmatrix} -1 \\ -5 \end{bmatrix}$; $W = \text{span}\left\{ \begin{bmatrix} -4 \\ 1 \end{bmatrix} \right\}$

(b) $\mathbf{v} = \begin{bmatrix} 0 \\ -8 \\ 6 \end{bmatrix}$; $W = \text{span}\left\{ \begin{bmatrix} -7 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 8 \\ -3 \\ -4 \end{bmatrix} \right\}$
[𝕊]-332

(c) $\mathbf{v} = \begin{bmatrix} -12 \\ 11 \\ -1 \end{bmatrix}$; $W = \text{span}\left\{ \begin{bmatrix} 5 \\ 11 \\ 61 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \\ -11 \end{bmatrix} \right\}$

(d) $\mathbf{v} = \begin{bmatrix} 7 \\ 7 \\ 12 \end{bmatrix}$; $W = \text{span}\left\{ \begin{bmatrix} 6 \\ 1 \\ -4 \end{bmatrix}, \begin{bmatrix} -4 \\ -7 \\ 6 \end{bmatrix} \right\}$
[𝔸]-361

6. For $\mathbf{v}$ and W in question 5, find the orthogonal decomposition of $\mathbf{v}$ relative to W. [𝕊]-333 [𝔸]-361

7. Solve the system. If it is inconsistent, find a best approximation to a solution.

(a) SageMathCell 98
$$\begin{array}{rcrcrcr} -6x & + & 3y & - & 15z & = & 7 \\ 14x & - & 7y & + & 35z & = & 1 \\ -18x & + & 9y & - & 45z & = & 6 \end{array}$$
[𝕊]-333

(b) SageMathCell 99
$$\begin{array}{rcrcrcr} 14x & + & 4y & + & 2z & = & 9 \\ 3x & - & y & + & 6z & = & 1 \\ -5x & - & 5y & + & 10z & = & 8 \end{array}$$

(c) SageMathCell 100
$$\begin{array}{rcrcrcr} 7x & + & 11y & + & 12z & = & -6 \\ 5x & - & 8y & - & 4z & = & -1 \\ & & 4y & - & 7z & = & -2 \end{array}$$
[𝔸]-361

(d) SageMathCell 101
$$\begin{array}{rcrcrcr} -4x & + & 2y & - & 10z & = & -11 \\ 8x & - & 4y & + & 20z & = & 6 \\ 2x & - & y & + & 5z & = & -7 \end{array}$$

(e) SageMathCell 102
$$\begin{array}{rcrcrcr} -3x & - & 6y & + & 3z & = & -12 \\ 3x & - & 2y & - & 19z & = & -12 \\ 2x & - & 11y & - & 32z & = & -37 \end{array}$$
[𝔸]-361

(f) SageMathCell 103
$$\begin{array}{rcrcrcr} 5x & - & 3y & + & z & = & -4 \\ -10x & + & 6y & - & 2z & = & 8 \\ 20x & - & 12y & + & 4z & = & -16 \end{array}$$

(g) SageMathCell 104
$$\begin{array}{rcrcrcr} x & - & 8y & + & 3z & = & 10 \\ -11x & + & 7y & - & 6z & = & 4 \\ -9x & + & 3y & - & 4z & = & 9 \end{array}$$
[𝔸]-361

8. Find a basis for $W^\perp$.

(a) $W = \text{span}\left\{ \begin{bmatrix} 10 \\ 5 \\ -4 \end{bmatrix} \right\}$

(b) $W = \text{span}\left\{ \begin{bmatrix} 7 \\ 11 \\ 12 \end{bmatrix} \right\}$ [𝕊]-333

(c) $W = \text{span}\left\{ \begin{bmatrix} -1 \\ -5 \\ 3 \end{bmatrix}, \begin{bmatrix} 9 \\ -8 \\ -1 \end{bmatrix} \right\}$

(d) $W = \text{span}\left\{ \begin{bmatrix} 5 \\ -3 \\ -9 \end{bmatrix}, \begin{bmatrix} -2 \\ -11 \\ 5 \end{bmatrix} \right\}$ [𝔸]-361

9. Argue that for any subspace W of an inner product space V, $\dim W + \dim W^\perp = \dim V$.

10. Argue that for any subspace W of an inner product space V, if $\mathbf{v}$ is in W, then $\text{proj}_W \mathbf{v} = \mathbf{v}$. [𝔸]-361

11. Argue that for any subspace W of an inner product space V, the orthogonal decomposition of $\mathbf{0}$ is $\mathbf{0} + \mathbf{0}$.

12. Let W be a subspace of an inner product space V, and let $\mathcal{B}$ be a basis of W. Argue that $\mathbf{v}$ is orthogonal to each vector in $\mathcal{B}$ if and only if $\mathbf{v}$ is in $W^\perp$. [𝔸]-361

The remainder of the exercises are set within the inner product space $\mathbb{P}_3(\mathbb{R})$ with inner product

$$\langle p, q \rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1) + p(2)q(2).$$

You may find the SageCell at SageMathCell 105 helpful. It computes this inner product for arbitrary third degree polynomials.

13. What multiple of $\mathbf{q}$ lands closest to $\mathbf{p}$ (is a best approximation)?

(a) $\mathbf{p} = 7x^2 - 10x - 5$
$\mathbf{q} = 3x^2 + 12x + 6$

(b) $\mathbf{p} = -10x^2 + 5x + 4$
$\mathbf{q} = 3x^3 + 8x^2 - x - 7$

(c) $\mathbf{p} = x^3 - 11x^2 - 9x + 10$
$\mathbf{q} = 12x^3 - x - 5$ [𝕊]-334

14. How far off is the best approximation in question 13? [𝕊]-334

15. Find the orthogonal projection of $\mathbf{p}$ onto span$\mathcal{B}$. $\mathcal{B}$ is an orthogonal set.

(a) $\mathbf{p} = x^3 + 3x^2 - 6x + 4$
$\mathcal{B} = \left\{ -6x^2 + 2x - 1 \right\}$ [𝔸]-361

(b) $\mathbf{p} = 8x^2 - x - 12$
$\mathcal{B} = \left\{ x^3 + 2x^2 - 2x + 8 \right\}$

(c) $\mathbf{p} = 2x^3 - 9x^2 + 9x - 6$
$\mathcal{B} = \left\{ 11x^2 + 3x - 1, -4x^3 + 8x^2 + 17x - 28 \right\}$
[𝔸]-361

(d) $\mathbf{p} = 11x^3 + 9x - 10$
$\mathcal{B} = \left\{x^3 - 9x^2 - x - 2, -4x^3 - 12x^2 + 3x + 47\right\}$

(e) $\mathbf{p} = -2x^3 - 5x^2 + 5x - 11$
$\mathcal{B} = \left\{5x^2 - 11x + 14, 4x^2 - 12x - 5\right\}$
[A]-361

16. Find the orthogonal projection of **p** onto span$\mathcal{B}$. $\mathcal{B}$ is not an orthogonal set.

(a) $\mathbf{p} = 2x^3 + x^2 - x - 3$
$\mathcal{B} = \left\{8x^3 + 11x^2 - 6x, 9x^3 - 7x - 9\right\}$

(b) $\mathbf{p} = x^3 - 5x^2 + 12x - 7$
$\mathcal{B} = \left\{-3x^2 + 3x - 1, -7x^3 + 3x + 11\right\}$
[A]-361

(c) $\mathbf{p} = -10x^3 + 10x^2 + 11x + 12$
$\mathcal{B} = \left\{9x^3 - 9x, -6x^2 - 1\right\}$

17. Is **p** in $W^\perp$?

(a) $\mathbf{p} = 5x^3 - x^2 + 3x + 2$
$W = \operatorname{span}\left\{7x^3 + 8x^2 - 8x - 1\right\}$ [A]-362

(b) $\mathbf{p} = 5x^3 - 8x^2 - 11x + 6$
$W = \operatorname{span}\left\{5x^3 + 3x^2 + 36, 4x^2 - x + 2\right\}$

(c) $\mathbf{p} = x^3 + 10x^2 - 2x - 4$
$W = \operatorname{span}\left\{-4x^3 + 3x^2 - 2x - 6, 3x^3 + 5x^2\right\}$

18. For **p** and W in question 17, find the orthogonal decomposition of **p** relative to W. [A]-362

19. Find a basis for $W^\perp$.

(a) $W = \operatorname{span}\left\{1, x, x^2 - 2\right\}$

(b) $W = \operatorname{span}\left\{x + 1, x^2 - 2x - 2\right\}$ [A]-362

Answers

minimum integer solution We know that $|(54x + 30y) - 17|$ cannot equal zero, so the best we can hope for is to find integers x and y so that $|(54x + 30y) - 17| = 1$. That is, $(54x + 30y) - 17 = 1$ or $(54x + 30y) - 17 = -1$. Adding 17 to both sides of these equations, we seek integer solutions of $54x + 30y = 18$ or $54x + 30y = 16$. Since 16 is not a multiple of 6, there is no hope of finding integer solutions of $54x + 30y = 16$. Since 18 is a multiple of 6, perhaps there are integer solutions of $54x + 30y = 18$. Dividing both sides of the equation by 6, $9x + 5y = 3$ or $9x = 3 - 5y$. As long as $3 - 5y$ is a multiple of 9, we will have a solution. For example, $y = -3$ makes $3 - 5y = 18$ (and $x = 2$ makes $9x = 18$), so one solution is $(x, y) = (2, -3)$. Sure enough, $|(54 \cdot 2 + 30 \cdot -3) - 17| = |108 - 90 - 17| = 1$. There are others.

inconsistent system The augmented matrix for the system reduces as follows.

$$\begin{bmatrix} 1 & -3 & 8 \\ 4 & 9 & 5 \\ -9 & 6 & 7 \end{bmatrix} \to \begin{bmatrix} 1 & -3 & 8 \\ 0 & 21 & -27 \\ 0 & -21 & 79 \end{bmatrix} \to \begin{bmatrix} 1 & -3 & 8 \\ 0 & 21 & -27 \\ 0 & 0 & 52 \end{bmatrix}$$

An echelon form has a pivot in the rightmost column (the third row represents the equation $0 = 52$), so the system is inconsistent.

$W^\perp$ is a subspace Let **u** and **v** be in $W^\perp$ and c be scalar. (By definition, $\langle \mathbf{u}, \mathbf{w}\rangle = \langle \mathbf{v}, \mathbf{w}\rangle = 0$ for all **w** in W). Then for any **w** in W,

$$\langle \mathbf{0}, \mathbf{w}\rangle = \langle 0\mathbf{w}, \mathbf{w}\rangle = 0\langle \mathbf{w}, \mathbf{w}\rangle = 0$$

and

$$\langle \mathbf{u} + \mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle = 0$$

and

$$\langle c\mathbf{u}, \mathbf{w}\rangle = c\langle \mathbf{u}, \mathbf{w}\rangle = c \cdot 0 = 0$$

so $\mathbf{0}$, $\mathbf{u} + \mathbf{v}$, and $c\mathbf{u}$ are all in $W^\perp$. Since $W^\perp$ is a subset of V, this is sufficient to show that $W^\perp$ is a subspace.

$\mathbf{v} - \mathbf{proj}_W \mathbf{v}$ is in $W^\perp$ Suppose **v** is orthogonal to every element of *any* basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ of a subspace W. Then for any scalars $c_1, c_2, \ldots, c_p$,

$$\begin{aligned}\langle \mathbf{v}, c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_p\mathbf{b}_p\rangle &= \langle \mathbf{v}, c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_p\mathbf{b}_p\rangle \\ &= \langle \mathbf{v}, c_1\mathbf{b}_1\rangle + \langle \mathbf{v}, c_2\mathbf{b}_2\rangle + \cdots + \langle \mathbf{v}, c_p\mathbf{b}_p\rangle \\ &= c_1\langle \mathbf{v}, \mathbf{b}_1\rangle + c_2\langle \mathbf{v}, \mathbf{b}_2\rangle + \cdots + c_p\langle \mathbf{v}, \mathbf{b}_p\rangle \\ &= 0.\end{aligned}$$

Since every vector in W has the form $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_p$, this shows $\mathbf{v}$ is orthogonal to every vector in W and therefore is in $W^\perp$.

REMARK: Note that if $\mathbf{v}$ is in $W^\perp$, then $\mathbf{v}$ is orthogonal to every element of *any* basis $\mathcal{B}$ (since $\mathbf{v}$ is orthogonal to *every* vector in W—including basis vectors). Altogether we have that $\mathbf{v}$ is in $W^\perp$ if and only if $\mathbf{v}$ is orthogonal to every element of a basis of W.

Pythagorean theorem Because $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ and therefore

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 . \end{aligned}$$

theorem 18 implies corollary 19 Corollary 19 is the special case of theorem 18 where W is the column space of M, $V = \mathbb{R}^m$ and $\mathbf{v} = \mathbf{b}$.

Chapter 7

Further Applications

7.1 Linear Regression [6.4]

Perhaps the most ubiquitous application of linear algebra outside the boundaries of mathematics is linear regression, used to test hypotheses and produce models of phenomena in innumerable fields including meteorology, criminology, economics, materials science, archaeology, enginering, and psychology [21, 8, 16, 6, 1, 23, 3]. Anywhere two or more quantities are suspected of correlation, regression analysis can be performed. In its simplest form, two quantities are suspected of having a linear relationship. Data are collected on the two quantities, and a model (linear function) predicting one quantity based on the other is produced and analyzed.

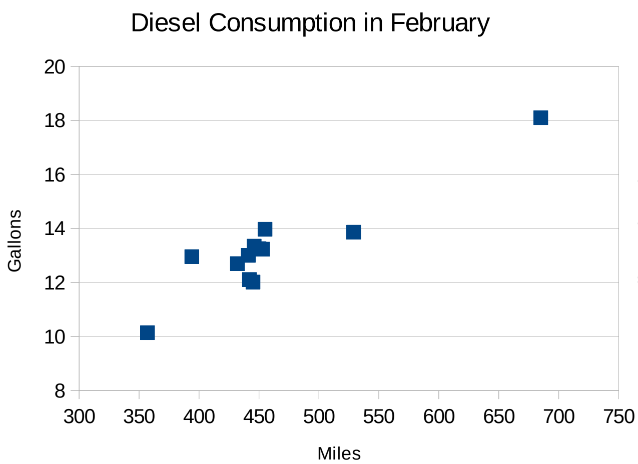
For example, it is well known that the distance a gas or diesel powered vehicle is driven (in miles, for example) is more or less directly proportional to the volume of fuel (in gallons, for example) consumed. Also understood is that highway driving generally uses less fuel per mile than city driving. This is why statistics on new cars will include both a highway and a city mileage estimate. The graphs in figure 7.1.1 were produced from the February driving data for a 2010 VW Jetta Sportwagen TDI in the chart below. Only February data are considered because it is also known that ambient temperature affects a combustion engine's efficiency. This car was driven in New England, where the average temperature in February is around 30°F, 0°C.

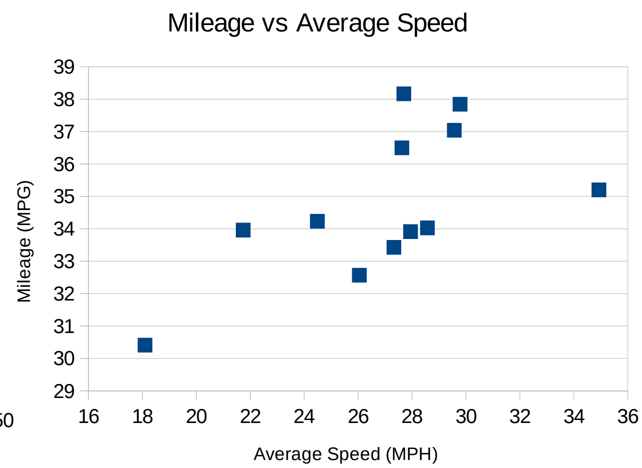
The graphs confirm the claims that driving longer distances requires more fuel (left graph) and that highway driving uses fuel more efficiently than city driving (as average speed increases so does mileage, right graph). When trends like these are observed, linear regression provides a way to quantify the relationship between the variables in the form of a function. This function can then be used to predict one quantity from the other.

Fill-up Date	Elapsed Miles	Gallons	Average Speed	Price per Gallon
02/08/12	450	13.25	21.739	4.36
02/23/12	685	18.101	29.783	4.40
02/17/13	394	12.956	18.098	4.36
02/01/14	445	12.014	29.568	4.38
02/16/14	432	12.696	28.571	4.60
02/26/14	529	13.861	27.696	4.46
02/06/15	453	13.233	24.486	3.25
02/22/15	357	10.142	34.932	3.00
02/12/16	442	12.11	27.625	2.30
02/16/17	455	13.971	26.045	2.68
02/02/18	446	13.343	27.328	3.10
02/20/18	441	13.003	27.947	3.15

Let's say the owner of this vehicle is planning a trip from New Haven, CT to Augusta, ME (approximately 600-miles round trip) next February and is interested in how much fuel will be used. Perhaps the simplest way to estimate

Figure 7.1.1: Graphs of Diesel Data





is to sum the elapsed miles, sum the gallons, and divide. This gives an average of about 0.0286996 gallons per mile. A 600-mile trip at this rate of consumption would require $0.0286996 \cdot 600 = 17.21976$ gallons of diesel.

For this application, that is probably good enough. However, we can do slightly better using linear regression. We know that fuel consumed is (roughly) directly proportional to miles driven, so they are related by a function of the form $y = kx$. Either variable can represent either quantity, but since we are interested in predicting fuel required given distance driven, we are looking for a function of the form $f(x) = kx$ where $x =$ distance driven and $f(x)$ is the fuel required. The simple average calculated above produces the model $f(x) = 0.0286996x$, but is this the best value for k?

It depends on how you define "best value", but one reasonable definition is to minimize the sum of the squared errors, where an error, also known as a residual, is the difference between an observed response and the modeled response to the same input. For example, 12.956 is the observed response to the input 394 (observed on 02/17/13). The modeled response, using $f(x) = 0.0286996x$, is $f(394) = 0.0286996 \cdot 394 \approx 11.308$ gallons, however. The error, or residual, for this observation is therefore $11.308 - 12.956$ or -1.648 gallons. The squared error is $(-1.648)^2 \approx 2.716$. A similar squared error can be calculated for each observation. The sum of the squared errors is, accurate to 5 decimal places, 9.30835.

As a linear algebra problem, finding the best value of k in this sense amounts to finding the best approximation (see corollary 19) of $M\mathbf{v} = \mathbf{b}$ where

$$M = \begin{bmatrix} 450 \\ 685 \\ 394 \\ \vdots \\ 441 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} k \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 13.25 \\ 18.101 \\ 12.956 \\ \vdots \\ 13.003 \end{bmatrix}$$

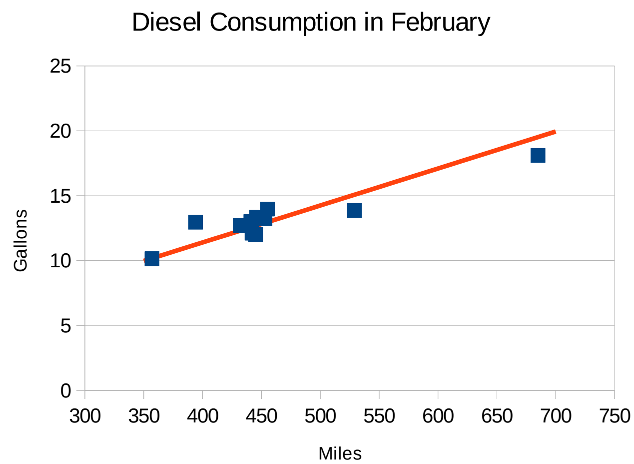
since

$$\begin{aligned} \|M\mathbf{v} - \mathbf{b}\| &= \left\| \begin{matrix} 450k - 13.25 \\ 685k - 18.101 \\ 394k - 12.956 \\ \vdots \\ 441k - 13.003 \end{matrix} \right\| \\ &= \sqrt{(450k - 13.25)^2 + \cdots + (441k - 13.003)^2}, \end{aligned}$$

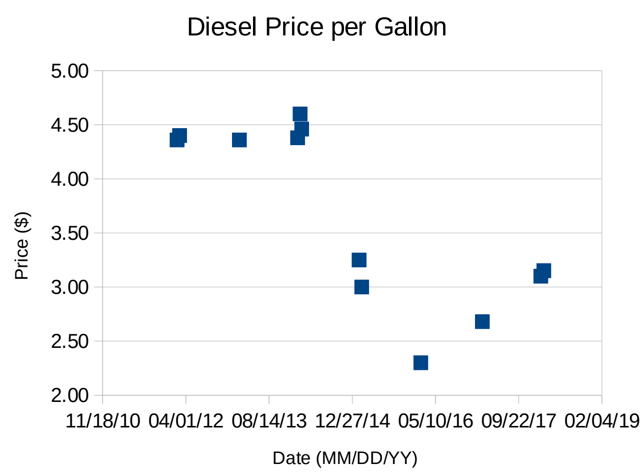
the square root of the sum of the squared errors. And 19 tells us the best approximation is the projection of $\mathbf{b}$ onto the column space of M. Letting W be the column space of M,

$$\begin{aligned}\mathrm{proj}_W\mathbf{b} = \mathrm{proj}_{M_{:,1}}\mathbf{b} &= \frac{\mathbf{b}\cdot M_{:,1}}{M_{:,1}\cdot M_{:,1}}M_{:,1} = \frac{74639.689}{2619895} \\ &\approx 0.0284896M_{:,1}\end{aligned} \tag{7.1.1}$$

so the best value of k is 0.0284896 (giving about 17.09 gallons for a 600-mile trip). The sum of the squared errors for the model $f(x) = 0.0284896x$ is $\left\|M\left[\begin{array}{c}0.0284896\end{array}\right] - \mathbf{b}\right\|^2 \approx 9.19278$—slightly lower than the 9.30835 we got from the model $f(x) = 0.0286996x$. Plotting the model on the same axes as the data illustrates the closeness of fit.



The driving data provide other opportunities for linear regression. Plotting the price of diesel over time produces the following graph.



This graph shows an overall downward trend in price over the six year span of the data. Linear regression can be used to capture this overall trend. If we are interested in an average decresase in price over this time span, we could find the best fit model of the form $p(t) = p_0 + rt$, a linear model whose slope estimates the average annual drop in price.

As a linear algebra problem, we wish to find the best approximation to $M\mathbf{v} = \mathbf{b}$ where M holds the inputs, $\mathbf{b}$ holds the responses, and $\mathbf{v}$ holds the unknown parameters p_0 and r. In this case, the input variable is time, which we will measure in days since 1 February 2012:

$$M = \begin{bmatrix} 1 & 7 \\ 1 & 22 \\ 1 & 382 \\ \vdots & \vdots \\ 1 & 2211 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} p_0 \\ r \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4.36 \\ 4.40 \\ 4.36 \\ \vdots \\ 3.15 \end{bmatrix},$$

which has best approximation

$$\begin{bmatrix} p_0 \\ r \end{bmatrix} \approx \begin{bmatrix} 4.55591498045673 \\ -0.000845069933663 \end{bmatrix}. \tag{7.1.2}$$

Since we measured time in days, r represents the average change in price per day, not year. To get an annual change, we multiply r by 365 to get -0.308, an average decrease of approximately 31 cents per year.

The graph of diesel price over time does not indicate a steady decline, however. While the overall trend is downward, there is a fluctuation as well. A more accurate model of the actual price over this time period would come from a model that caputures this fluctuation. For example, a model of the form $f(t) = p_0 + rt + \alpha\sin(\omega t) + \beta\cos(\omega t)$ might provide reasonable results since it includes a linear portion ($p_0 + rt$) to capture the overall decrease and periodic portion ($\alpha\sin(\omega t) + \beta\cos(\omega t)$) to capture the fluctuation. However, linear regression only approximates parameters that vary linearly with respect to the response and $f(t) = p_0 + rt + \alpha\sin(\omega t) + \beta\cos(\omega t)$ does not vary linearly with respect to its parameters.

Crumpet 30: Linear Variation

If the value of a function for each set of fixed inputs is a linear combination of its parameters, we say the function varies linearly with respect to its parameters or is linear in its parameters. Otherwise the function is nonlinear in its parameters.

For example, $f(1) = p_0 + r + \alpha\sin(\omega) + \beta\cos(\omega)$ is not a linear combination of p_0, r, α, β, and ω. It does not take the form $a_0p_0 + a_1r + a_2\alpha + a_3\beta + a_4\omega$. However, after fixing $\omega = \frac{2\pi}{5\times365}$, ω is no longer a parameter and $f(1) = p_0 + r + \sin\left(\frac{2\pi}{5\times365}\right)\cdot\alpha + \cos\left(\frac{2\pi}{5\times365}\right)\cdot\beta$ is a linear combination of its parameters p_0, r, α, β, and f is so for every other value of t too. Setting the value $\omega = \frac{2\pi}{5\times365}$ gives the sine and cosine functions a 5-year period.

As a linear algebra problem, we wish to find the best approximation to $M\mathbf{v} = \mathbf{b}$ where M holds the inputs, $\mathbf{b}$ holds the responses, and $\mathbf{v}$ holds the unknown parameters. Again, time will be measured in days since 1 February 2012.

$$M = \begin{array}{c} \begin{array}{cccc} & t & \sin(\omega t) & \cos(\omega t) \end{array} \\ \begin{bmatrix} 1 & 7 & 0.0172134 & 0.999852 \\ 1 & 22 & 0.0540754 & 0.998537 \\ 1 & 382 & 0.807206 & 0.590269 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 2211 & -0.748605 & 0.663016 \end{bmatrix} \end{array}, \quad \mathbf{v} = \begin{bmatrix} p_0 \\ r \\ \alpha \\ \beta \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4.36 \\ 4.40 \\ 4.36 \\ \vdots \\ 3.15 \end{bmatrix},$$

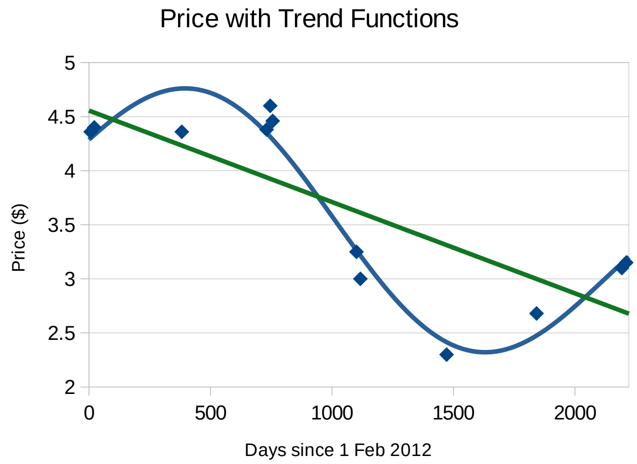
which has best approximation

$$\begin{bmatrix} p_0 \\ r \\ \alpha \\ \beta \end{bmatrix} \approx \begin{bmatrix} 4.47274711624545 \\ -0.000907047402863 \\ 0.71642901672004 \\ -0.22510779212064 \end{bmatrix}. \tag{7.1.3}$$

The sums of the squared errors for the two models

$$p(x) = 4.55591 - 845070(10)^{-4}t$$
$$f(x) = 4.47275 - 907047(10)^{-4}t + 0.716429\sin(\omega t) - 0.225108\cos(\omega t)$$

are 3.02939 and 0.460556, respectively. The graphs of the two models superimposed on the data clearly illustrate how much closer f comes to predicting the observed data.



For a final linear regression on the February driving data, we return to the thought that mileage is affected by driving speed. One model that incorporates this fact is $g(x,s) = (k+\beta s)x = kx + \beta sx$, where x is distance (as before) and s is average speed. The number of gallons consumed per mile, $(k+\beta s)$, varies with average speed s and $g(0,s) = 0$. This model is slightly different from the ones we have derived so far. Here, we have two input variables, making this a **multiple linear regression**, or **multilinear regression**. The principle is the same, however. We wish to find the best approximation of $M\mathbf{v} = \mathbf{b}$ where M holds the inputs, $\mathbf{b}$ holds the responses, and $\mathbf{v}$ holds the unknown parameters. In this case,

$$M = \begin{array}{c} \begin{array}{cc} \quad x \quad & \quad sx \quad \end{array} \\ \begin{bmatrix} 450 & 9782.55 \\ 685 & 20401.355 \\ 394 & 7130.612 \\ \vdots & \vdots \\ 441 & 12324.627 \end{bmatrix} \end{array}, \quad \mathbf{v} = \begin{bmatrix} k \\ \beta \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 13.25 \\ 18.101 \\ 12.956 \\ \vdots \\ 13.003 \end{bmatrix},$$

which has best approximation

$$\begin{bmatrix} k \\ \beta \end{bmatrix} \approx \begin{bmatrix} 0.0379147381262610 \\ -0.000346513745672586 \end{bmatrix} \tag{7.1.4}$$

and sum of squared errors 5.21664. Compare this to the 9.30835 we found without considering average driving speed.

Given that the hypothetical trip from New Haven to Augusta is to be driven mostly on the highway, we can approximate the required fuel by, for example,

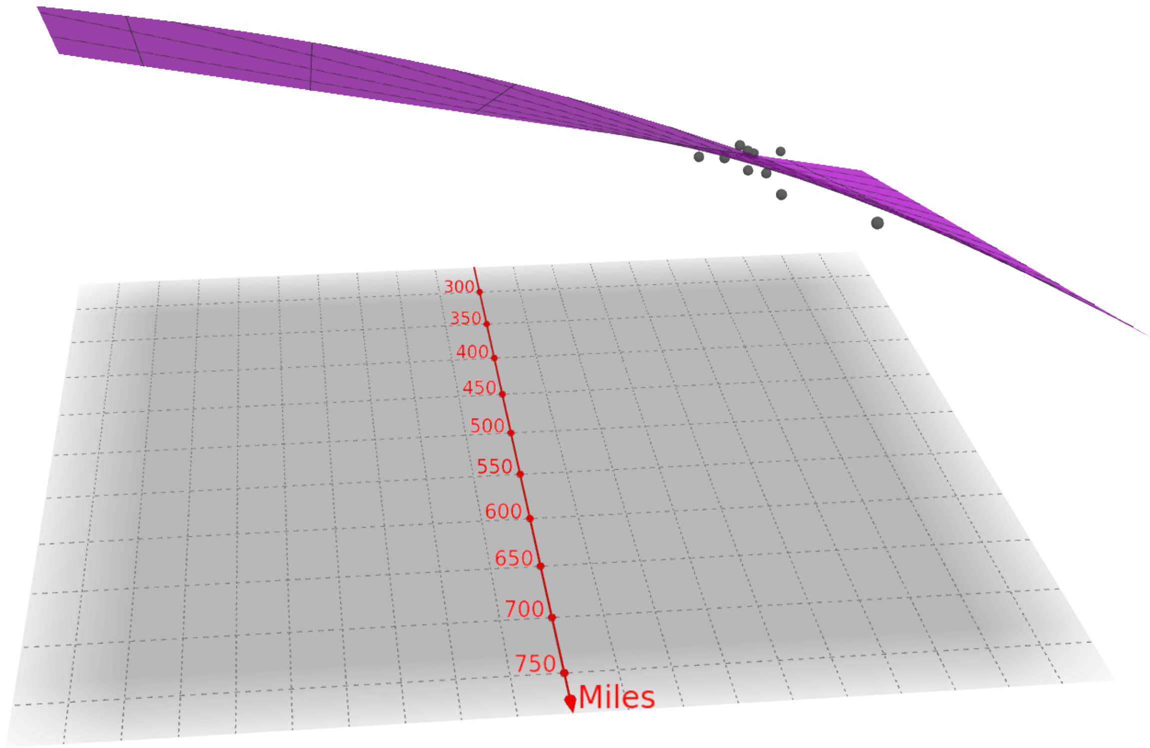
$$\begin{aligned} g(600,45) &= (0.0379147 - 0.000346514(45))600 \\ &\approx 13.39, \end{aligned}$$

significantly different from the original estimates of over 17 gallons. The 13.39 gallon estimate should be met with some skepticism, however. It uses an average speed of 45 miles per hour while the highest average speed for which we have data is about 35 miles per hour. There is no evidence that the model applies to an average speed of 45 miles per hour. More data and possibly a revision to the model should be considered before using an average speed of 45. On the other hand, hypothesizing that it is reasonable to expect the car's efficiency to be better at an average speed of 45 miles per hour than it is at an average speed of 35 miles per hour, we can use the model with an average speed of 35 to get an (expected) overestimate of the required volume of fuel.

$$\begin{aligned} g(600,35) &= (0.0379147 - 0.000346514(35))600 \\ &\approx 15.47 \end{aligned}$$

is still considerably less than 17—and likely an overestimate.

A linear regression model with two input parameters is, geometrically, a regression surface. A plot of $g(x,s)$ with the twelve data points is shown below.



Normal Equations

As presented in section 6.4, the calculation of a best approximation involves projecting onto the column space of a coefficient matrix, requiring an orthogonal basis for the column space. While (Gram-Schmidt) orthogonalization can be applied to find such a basis, the process is computationally intensive and, more detrimental to the results, error prone. In practice, the normal equations

$$M^T M\mathbf{v} = M^T\mathbf{b}$$

are solved instead. It is known that $\mathbf{v}$ is a solution of the normal equations if and only if $\mathbf{v}$ is a best approximation to a solution of $M\mathbf{v} = \mathbf{b}$.

Crumpet 31: Normal Equations

Letting W be the column space of M, an $m \times n$ matrix, the following statements are equivalent.

1. $\hat{\mathbf{v}}$ is a best approximation to a solution of $M\mathbf{v} = \mathbf{b}$.
2. $M\hat{\mathbf{v}}$ is the closest point to $\mathbf{b}$ in the column space of M.
3. $M\hat{\mathbf{v}} = \mathrm{proj}_W\mathbf{b}$.
4. $\mathbf{b} - M\hat{\mathbf{v}}$ is in $W^\perp$.
5. $(\mathbf{b} - M\hat{\mathbf{v}})^T M_{:,j} = 0$ for all $j = 1, 2, \ldots, n$.
6. $M^T M\hat{\mathbf{v}} = M^T\mathbf{b}$.

$1 \Leftrightarrow 2$ by definition of best approximation. $2 \Leftrightarrow 3$ by theorem 19. $3 \Rightarrow 4$ by the fact that $(\mathbf{b} - \mathrm{proj}_W\mathbf{b})$ is in $W^\perp$. $4 \Rightarrow 3$ by the facts that (i) $\mathbf{b} = M\hat{\mathbf{v}} + (\mathbf{b} - M\hat{\mathbf{v}})$; (ii) $M\mathbf{v}$ is in W; (iii) $\mathbf{b} - M\hat{\mathbf{v}}$ is in $W^\perp$; and (iv) corollary 17. $4 \Leftrightarrow 5$ by definition of $W^\perp$ and the fact that each column of M is in W. $5 \Leftrightarrow 6$ by matrix algebra: for each $j = 1, 2, \ldots, n$, $(\mathbf{b} - M\hat{\mathbf{v}})^T M_{:,j} = 0 \Leftrightarrow M_{:,j}^T(\mathbf{b} - M\hat{\mathbf{v}}) = 0 \Leftrightarrow M_{:,j}^T\mathbf{b} - M_{:,j}^T M\hat{\mathbf{v}} = 0 \Leftrightarrow M_{:,j}^T\mathbf{b} = M_{:,j}^T M\hat{\mathbf{v}}$, this last equality being true for all j if and only if $M^T M\hat{\mathbf{v}} = M^T\mathbf{b}$.

Because the set of best approximations of $M\mathbf{v} = \mathbf{b}$ equals precisely the solution set of $M^T M\hat{\mathbf{v}} = M^T\mathbf{b}$, the linear system $M\mathbf{v} = \mathbf{b}$ has a unique best approximation for each $\mathbf{b}$ in $\mathbb{R}^m$ if and only if $M^T M\hat{\mathbf{v}} = M^T\mathbf{b}$ has a unique solution for each $\mathbf{b}$ in $\mathbb{R}^m$. By theorem 7 $M^T M\hat{\mathbf{v}} = M^T\mathbf{b}$ has a unique solution for each $\mathbf{b}$ in $\mathbb{R}^m$ if and only if $M^T M\hat{\mathbf{v}} = \mathbf{0}$ has only the trivial solution if and only if $M^T M$ is invertible. Hence the linear system $M\mathbf{v} = \mathbf{b}$ has a unique best approximation for each $\mathbf{b}$ in $\mathbb{R}^m$ if and only if $M^T M$ is invertible.

Solving the normal equations amounts to solving a linear system of p equations in p variables where p is the number of parameters (not the number of data points). The normal equations represent a relatively small system with known, dependable solution techniques.

For example, the model $p(t) = p_0 + rt$, which came from a best approximation with

$$M = \begin{bmatrix} 1 & 7 \\ 1 & 22 \\ 1 & 382 \\ \vdots & \vdots \\ 1 & 2211 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} p_0 \\ r \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4.36 \\ 4.40 \\ 4.36 \\ \vdots \\ 3.15 \end{bmatrix}$$

can be solved by first computing

$$M^T M = \begin{bmatrix} 12 & 12580 \\ 12580 & 19526278 \end{bmatrix} \quad \text{and} \quad M^T\mathbf{b} = \begin{bmatrix} 44.04 \\ 40812.34 \end{bmatrix}$$

and then solving

$$\begin{bmatrix} 12 & 12580 \\ 12580 & 19526278 \end{bmatrix}\begin{bmatrix} p_0 \\ r \end{bmatrix} = \begin{bmatrix} 44.04 \\ 40812.34 \end{bmatrix}.$$

Can you provide this solution? Answer on page 236.

Key Concepts

least squares solution a best approximation $\hat{\mathbf{v}}$ of $M\mathbf{v} = \mathbf{b}$, having sum of squared errors $||M\hat{\mathbf{v}} - \mathbf{b}||$.

sum of squared errors given observations (X_i, y_i), $i = 1, 2, \ldots, N$, and model $y = f(X)$, $(f(X_1) - y_1)^2 + (f(X_2) - y_2)^2 + \cdots + (f(X_N) - y_N)^2$.

linear regression given a model of the form $f(X) = \beta_1 f_1(X) + \beta_2 f_2(X) + \cdots + \beta_p f_p(X)$ and observations (X_i, y_i), $i = 1, 2, \ldots, N$, linear regression refers to finding a best approximation of

$$\begin{bmatrix} f_1(X_1) & f_2(X_1) & \cdots & f_p(X_1) \\ f_1(X_2) & f_2(X_2) & \cdots & f_p(X_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(X_N) & f_2(X_N) & \cdots & f_p(X_N) \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}.$$

normal equations $M^T M\mathbf{v} = M^T\mathbf{b}$, whose solutions coincide precisely with best approximations of $M\mathbf{v} = \mathbf{b}$.

multiple linear regression linear regression with multiple input variables.

multilinear regression another name for multiple linear regression.

Exercises

1. Is the function linear or nonlinear in its parameters? A parameter is any variable quantity not listed as an independent variable (input) of the function.

   (a) $f(x) = a_3x^3 + a_2x^2 + a_1x + a_0$

   (b) $g(x) = (x - r_1)(x - r_2)(x - r_3)$ [𝔸]-362

   (c) $h(t) = P_0e^{rt}$

   (d) $m(t) = \frac{k}{t} + r\ln(2\pi t)$ [𝔸]-362

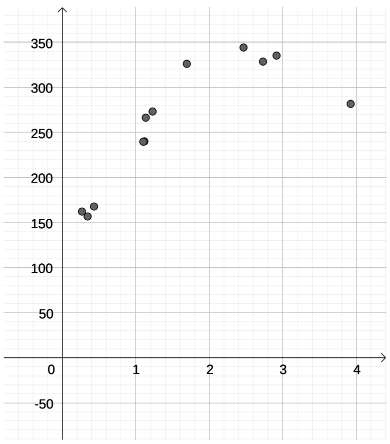
   (e) $\Phi_E(Q) = \dfrac{Q}{\varepsilon_0}$ [Gauss's Law of Electric Flux]

(f) $G(m_1, m_2, r) = \dfrac{km_1m_2}{r^2}$ [Gravitational Force] [𝔸]-362

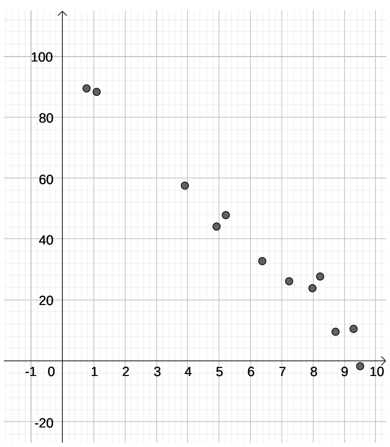
(g) $F(m, \theta) = \mu mg\cos\theta$ [Frictional Force] where g is a gravitational constant, not a parameter.

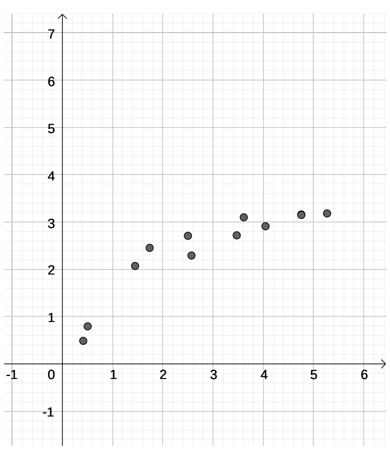
(h) $Y(A, L, \Delta_L) = \frac{FL}{A\Delta_L}$ [Modulus of Elasticity] [𝔸]-362

2. Based on the general shape of the graph, propose a model with one independent variable that could be fitted to the data using linear regression.

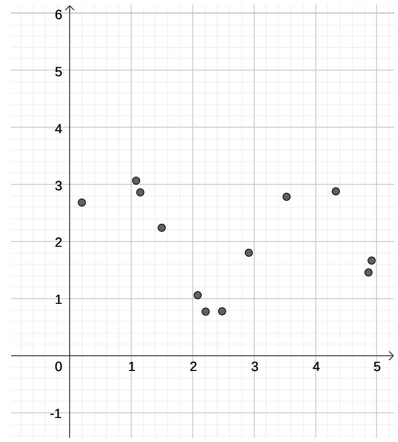
(a) 

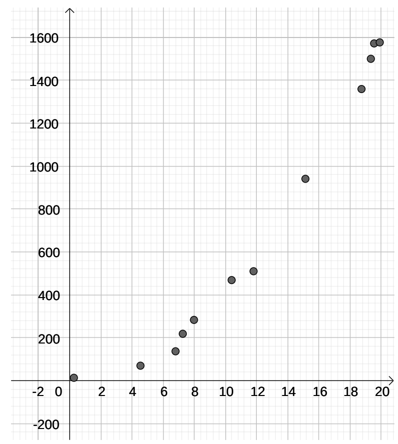
[𝔸]-362

(b) 

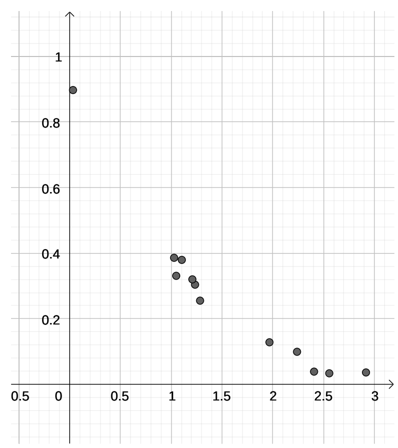
(c) 

[𝔸]-362

(d) 

(e) 

[𝔸]-362

(f) 

3. Use normal equations to find the best fit of the model to the data.

(a) SageMathCell 106

$f(x) = \beta_3x^3 + \beta_2x^2 + \beta_1x + \beta_0$

x	.389	.851	2.467	4.113
$f(x)$	141.1	−35.51	167.1	18.3

x	4.525	6.639	8.873	11.24
$f(x)$	173	243	1039	1783

[𝕊]-334

(b) SageMathCell 107

$g(x) = \beta_0 + \beta_1 x$

x	.94	3.002	4.837	7.422
$g(x)$	6.341	19.43	36.54	53.86

x	8.038	10.06	13.06	13.89
$g(x)$	47.57	61.6	86.13	83.31

[A]-362

(c) SageMathCell 108

$h(x) = \beta_0 + \beta_1 \sin(x) + \beta_2 \cos(x)$

x	.17	.275	.525	3.185
$h(x)$	10.79	10	8.533	−2.009

x	3.545	4.618	6.604	6.679
$h(x)$	−.437	7.455	10.02	9.357

(d) SageMathCell 109

$j(t) = \beta_0 \ln(3t)$

t	.042	1.778	1.934	3.431
$j(t)$	−13.22	10.7	11.18	14.8

t	3.888	5.491	6.57	8.98
$j(t)$	15.52	17.39	18.7	20.67

(e) SageMathCell 110

$k(t, x) = \beta_0 x^2 + \beta_1 xt + \beta_2 t^2$

$k(t, x)$ ← x →, ↑ t ↓

t \ x	**.041**	**.306**	**1.07**	**1.92**
.626	−1.36	8.17	6.59	24.1
1.01	8.47	8.88	16.9	38
1.53	18.9	19	29	46.1
1.63	21.4	28.9	36	50

[S]-335

(f) SageMathCell 111

$\ell(t, x) = \dfrac{\beta_0 x}{1 + e^t}$

$k(t, x)$ ← x →, ↑ t ↓

t \ x	**.53**	**1.27**	**1.95**	**2.15**
.14	5.31	10.5	17.6	37.9
.8	.41	6.39	21.4	44.1
1.51	11.4	22.6	39.4	60.1
2.24	40.4	39.4	56.6	80.8

[A]-362

(g) SageMathCell 112

$m(t, x) = \beta_0 + \frac{\beta_1}{1+x^2+e^{-t}}$

$m(t, x)$ ← x →, ↑ t ↓

t \ x	**.8**	**1.55**	**1.82**	**2.02**
.64	4.3	8.41	26.4	35.8
.83	7.04	9.67	25.3	43.5
1.49	16.7	19.2	38.8	58.5
1.85	26.6	31.9	50.9	69

(h) SageMathCell 113

$n(x_1, x_2) = \beta_0 \sin(x_1 x_2) + \beta_1 \cos(x_1 + x_2)$

$n(x_1, x_2)$ ← x_2 →, ↑ x_1 ↓

x_1 \ x_2	**.9**	**1.7**	**1.95**	**2.38**
.76	1.43	8.5	27	41.3
1.22	9.46	18.3	33.6	49
2.05	31.8	42.2	48.6	67.5
2.1	29.3	42.6	57.3	69.6

4. Redo questions 3abdef using orthogonal projection instead. [S]-336 [A]-362

5. Calculate the sum of the squared errors for the model of question 3. [S]-338 [A]-362

6. Fitting an exponential function. Many physical quantities are related exponentially, making $y(t) = ae^{kt}$ a common model in science. Unfortunately this model is not linear in its parameters, making linear regression impossible directly. By taking the natural log of both sides of the formula, however, the model becomes

$$\ln(y(t)) = \ln(a) + kt$$

which is a line in the variables t and $\ln y$. Find a best-fit exponential model for the data by following the outlined steps. [A]-362

t	.6203	1.062	1.625	2.158
$y(t)$	25.90	21.77	18.38	14.64

t	3.147	8.259	8.931	9.519
$y(t)$	9.905	2.818	3.022	2.110

(a) Complete the following chart, filling in the logarithms of the given y values.

t	.6203	1.062	1.625	2.158
$\ln y$	3.524	3.081	2.911	

t	3.147	8.259	8.931	9.519
$\ln y$				

(b) Fit a linear model, $f(t) = \beta_0 + \beta_1 t$ to the data in the chart of part (a).

(c) $\beta_0 = \ln(a)$ and $\beta_1 = k$ so the model is

$$y(t) = e^{\beta_0} e^{\beta_1 t}$$

Calculate $a = e^{\beta_0}$. You now have the parameters a and k for the model.

(d) Plot the model superimposed upon a scatterplot of the data to see the fit.

7. Eyeball challenge.

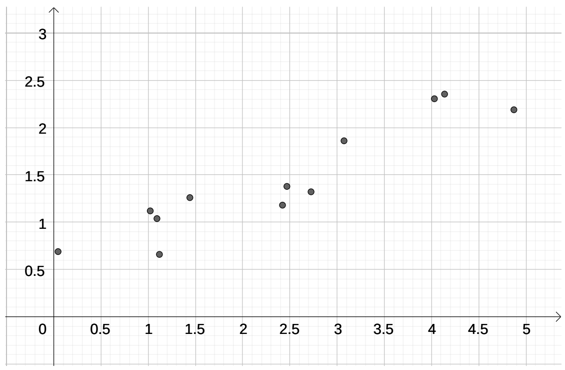
(a) Draw a line on the graph that fits the data well.

(b) Find an equation for the line you have drawn.

(c) Calculate the sum of the squared errors of your model (linear equation).

(d) Calculate the linear regression model (of the form $f(x) = \beta_0 + \beta_1 x$) for the data.

(e) Calculate the sum of the squared errors for the linear regression model. Compare it to your answer for 7c. How did you do with your "eyeballed" line?



The coordinates of the points are $(4.13, 2.35)$, $(1.02, 1.12)$, $(2.46, 1.37)$, $(.045, .688)$, $(1.09, 1.03)$, $(2.42, 1.17)$, $(1.11, .659)$, $(4.03, 2.3)$, $(2.72, 1.32)$, $(1.44, 1.25)$, $(4.87, 2.18)$, and $(3.07, 1.86)$.

8. 114 Verify the result in (7.1.1) using normal equations.

9. 115 Verify the result in (7.1.3) using normal equations.

10. 116 Verify the result in (7.1.4) using normal equations.

Download complete data for the diesel mileage of the Sportwagen TDI referred to in this section at the ancillary website, section 7.1, to complete the following exercises. Use data from June (or some other month) instead of data from February.

11. Recompute the model of (7.1.1) using data from the month of your choosing.

12. Recompute the model of (7.1.2) using data from the month of your choosing.

13. Recompute the model of (7.1.3) using data from the month of your choosing.

14. Recompute the model of (7.1.4) using data from the month of your choosing.

15. Create your own question, propose a model to help answer it, and find the parameters of your model using linear regression.

Answers

linear system solution The solution can be reached by row reduction or the inverse method. Since the coefficient matrix is 2×2 and it is easy enough to compute a 2×2 inverse, that is probably the easiest route:

$$\begin{aligned}
\begin{bmatrix} 12 & 12580 \\ 12580 & 19526278 \end{bmatrix}^{-1} &= \frac{1}{12(19526278) - 12580^2} \begin{bmatrix} 19526278 & -12580 \\ -12580 & 12 \end{bmatrix} \\
&= \begin{bmatrix} \frac{19526278}{76058936} & -\frac{12580}{76058936} \\ -\frac{12580}{76058936} & \frac{12}{76058936} \end{bmatrix} \approx \begin{bmatrix} 0.256726 & -1.65398(10)^{-4} \\ -1.65398(10)^{-4} & 1.57772(10)^{-7} \end{bmatrix}
\end{aligned}$$

so

$$\begin{aligned}
\mathbf{v} = \begin{bmatrix} p_0 \\ r \end{bmatrix} &\approx \begin{bmatrix} 0.256726 & -1.65398(10)^{-4} \\ -1.65398(10)^{-4} & 1.57772(10)^{-7} \end{bmatrix} \begin{bmatrix} 44.04 \\ 40812.34 \end{bmatrix} \\
&\approx \begin{bmatrix} 4.556 \\ -8.451(10)^{-4} \end{bmatrix},
\end{aligned}$$

the same as result (7.1.2).

7.2 Markov Chains [3.5]

From https://www.capitalbikeshare.com/about

> Capital Bikeshare is metro DC's bikeshare system, with more than 4,300 bikes available at 500 stations across six jurisdictions: Washington, DC; Arlington, VA; Alexandria, VA; Montgomery County, MD; Prince George's County, MD; Fairfax County, VA; and the City of Falls Church, VA. Capital Bikeshare provides residents and visitors with a convenient, fun and affordable transportation option for getting from Point A to Point B.
>
> Capital Bikeshare, like other bikeshare systems, consists of a fleet of specially designed, sturdy and durable bikes that are locked into a network of docking stations throughout the region. The bikes can be unlocked from any station and returned to any station in the system, making them ideal for one-way trips. People use bikeshare to commute to work or school, run errands, get to appointments or social engagements and more.
>
> Capital Bikeshare is available for use 24 hours a day, 7 days a week, 365 days a year. Riders have access to a bike at any station across the system.

Capital Bikeshare makes their trip data available to the general public free of charge. The data includes (i) duration of trip, (ii) start date and time, (iii) end date and time, (iv) starting station name and number, (v) ending station name and number, and more.[1]

The following chart was processed from real Capital Bikeshare data for the year 2018. It shows the total number of rides that started and ended in the section of Alexandria containing bike stations 31041 through 31048. The locations of the stations are to the right of the chart. The total number of rides accounted for is 10, 364.

		From								
		31041	**31042**	**31043**	**31044**	**31045**	**31046**	**31047**	**31048**	
To	**31041**	664	163	77	99	103	55	265	256	**Prince St & Union St**
	31042	124	519	152	159	161	101	658	710	**Market Square / King St & Royal St**
	31043	66	206	102	87	58	129	611	18	**Saint Asaph St & Pendleton St**
	31044	56	121	55	156	32	73	114	501	**King St & Patrick St**
	31045	41	128	41	22	70	121	64	187	**Commerce St & Fayette St**
	31046	20	97	98	41	76	96	70	11	**Henry St & Pendleton St**
	31047	172	561	568	88	64	180	182	28	**Braddock Rd Metro**
	31048	78	215	17	197	41	13	33	93	**King St Metro South**
	Total:	**1221**	**2010**	**1110**	**849**	**605**	**768**	**1997**	**1804**	**10364**

From these data, linear algebra can be applied to estimate the distribution of bicycles among the stations. Such information could be used to decide which stations to expand or reduce, where another station might be needed, the most likely place to find a free bike, and so on.

The method is one of prediction over time based on percentages. Given the distribution of bikes among the stations (as percentages) at some time, it uses the data to predict the distribution of bikes at some fixed amount of time later. Assuming the annual data on bicycle movement is reflected monthly, we will use one month for the time step.

Because the method works with percentages, we do not need to know how many bikes are in the neighborhood. It could be 25 or 250. No matter. We begin by dividing each column of the chart by the total number of rides in that column. The first column is divided by 1221, the second column by 2010, and so on, resulting in a table whose columns all sum to 1. Accurate to five decimal places, this normalized chart is collected in the matrix M:

$$M = \begin{bmatrix} 0.54382 & 0.08109 & 0.06937 & 0.11661 & 0.17025 & 0.07161 & 0.13270 & 0.14191 \\ 0.10156 & 0.25821 & 0.13694 & 0.18728 & 0.26612 & 0.13151 & 0.32949 & 0.39357 \\ 0.05405 & 0.10249 & 0.09189 & 0.10247 & 0.09587 & 0.16797 & 0.30596 & 0.00998 \\ 0.04586 & 0.06020 & 0.04955 & 0.18375 & 0.05289 & 0.09505 & 0.05709 & 0.27772 \\ 0.03358 & 0.06368 & 0.03694 & 0.02591 & 0.11570 & 0.15755 & 0.03205 & 0.10366 \\ 0.01638 & 0.04826 & 0.08829 & 0.04829 & 0.12562 & 0.12500 & 0.03505 & 0.00610 \\ 0.14087 & 0.27910 & 0.51171 & 0.10365 & 0.10579 & 0.23438 & 0.09114 & 0.01552 \\ 0.06388 & 0.10697 & 0.01532 & 0.23204 & 0.06777 & 0.01693 & 0.01652 & 0.05155 \end{bmatrix}$$

[1] See https://www.capitalbikeshare.com/system-data.

M thereby represents the percentage of rides starting at the station represented by the column that end at the station represented by the row **one** month later. For example, the 0.06937 in the first row, third column means that 6.937% of the bikes at station number 31043 will be at station number 31041 in a month. Empirically speaking, about 7% of the bikes at station number 31043 are destined for station number 31041 over the course of the month. We know this, in fact, was the case over the whole of 2018, and we are assuming it makes a good estimate of the monthly migration of bikes within the neighborhood.

Multiplying $M_{2,1}M_{1,5}$ then represents the percentage of bikes at the station of column 5 (31045) that head for the station of column 1 (31041) first and then to the station of column 2 (31042). In other words, it is the percentage of bikes at station 31045 that can be expected to be at station 31042 **two** months later. Similarly, $M_{2,2}M_{2,5}$ represents the percentage of bikes at station 31045 that are destined for station 31042 via station 31042 (the second ride is from station 31042 back to station 31042) over the course of two months; $M_{2,3}M_{3,5}$ represents the percentage of bikes at station 31045 that are destined for station 31042 after being dropped at station 31043; and so on. The sum $M_{2,1}M_{1,5}+M_{2,2}M_{2,5}+\cdots+M_{2,8}M_{8,5}$ therefore represents the total percentage of the bikes at station 31045 that can be expected to be at station 31042 *after two months*. Notice that sum is just a row-column product (row 2 times column 5), which is the 2,5-entry of M^2.

Generalizing, the i,j-entry of

$$M^2 = \begin{bmatrix} 0.34772 & 0.14568 & 0.14367 & 0.16169 & 0.17933 & 0.14162 & 0.14917 & 0.16962 \\ 0.18009 & 0.25757 & 0.26034 & 0.24737 & 0.21787 & 0.22429 & 0.20081 & 0.22320 \\ 0.09918 & 0.14711 & 0.20639 & 0.09837 & 0.11594 & 0.15050 & 0.11192 & 0.09361 \\ 0.07128 & 0.08900 & 0.06889 & 0.13178 & 0.08122 & 0.07528 & 0.06639 & 0.10299 \\ 0.04552 & 0.05952 & 0.05190 & 0.06237 & 0.07117 & 0.06664 & 0.05208 & 0.05619 \\ 0.03239 & 0.05021 & 0.05196 & 0.04321 & 0.06101 & 0.07067 & 0.05955 & 0.05025 \\ 0.15859 & 0.18732 & 0.17162 & 0.16729 & 0.20517 & 0.21018 & 0.29330 & 0.17834 \\ 0.06525 & 0.06360 & 0.04523 & 0.08794 & 0.06829 & 0.06081 & 0.06678 & 0.12580 \end{bmatrix}$$

holds the percentage of bikes starting at station $3104j$ that end up at station $3104i$ after two months. Likewise, the i,j-entry of M^k holds the percentage of bikes starting at station $3104j$ that end up at station $3104i$ after k months. Multiplying M^2 by itself,

$$M^4 = \begin{bmatrix} 0.22039 & 0.18022 & 0.17861 & 0.18487 & 0.18748 & 0.17934 & 0.18079 & 0.18673 \\ 0.21605 & 0.22893 & 0.23122 & 0.22754 & 0.22432 & 0.22723 & 0.22265 & 0.22529 \\ 0.12247 & 0.13283 & 0.13920 & 0.12638 & 0.12854 & 0.13309 & 0.12804 & 0.12492 \\ 0.08042 & 0.08277 & 0.08089 & 0.08616 & 0.08189 & 0.08129 & 0.08003 & 0.08512 \\ 0.05346 & 0.05606 & 0.05568 & 0.05638 & 0.05587 & 0.05612 & 0.05539 & 0.05599 \\ 0.04633 & 0.05067 & 0.05076 & 0.04970 & 0.05058 & 0.05155 & 0.05181 & 0.04994 \\ 0.19212 & 0.20053 & 0.19847 & 0.19753 & 0.20252 & 0.20392 & 0.21272 & 0.19884 \\ 0.06877 & 0.06799 & 0.06518 & 0.07144 & 0.06880 & 0.06746 & 0.06857 & 0.07318 \end{bmatrix}$$

and then M^4 by itself,

$$M^8 = \begin{bmatrix} 0.19015 & 0.18855 & 0.18844 & 0.18879 & 0.18885 & 0.18851 & 0.18857 & 0.18887 \\ 0.22448 & 0.22495 & 0.22501 & 0.22487 & 0.22483 & 0.22494 & 0.22486 & 0.22484 \\ 0.12913 & 0.12955 & 0.12965 & 0.12942 & 0.12944 & 0.12956 & 0.12947 & 0.12938 \\ 0.08181 & 0.08186 & 0.08185 & 0.08189 & 0.08185 & 0.08185 & 0.08183 & 0.08189 \\ 0.05532 & 0.05542 & 0.05542 & 0.05541 & 0.05540 & 0.05542 & 0.05541 & 0.05541 \\ 0.04985 & 0.05004 & 0.05005 & 0.05001 & 0.05001 & 0.05005 & 0.05005 & 0.05000 \\ 0.20067 & 0.20111 & 0.20109 & 0.20102 & 0.20108 & 0.20116 & 0.20127 & 0.20103 \\ 0.06858 & 0.06853 & 0.06849 & 0.06858 & 0.06855 & 0.06852 & 0.06854 & 0.06859 \end{bmatrix}$$

and so on,

$$M^{16} = \begin{bmatrix} 0.18890 & 0.18890 & 0.18890 & 0.18890 & 0.18890 & 0.18890 & 0.18890 & 0.18890 \\ 0.22483 & 0.22483 & 0.22483 & 0.22483 & 0.22483 & 0.22483 & 0.22483 & 0.22483 \\ 0.12944 & 0.12944 & 0.12944 & 0.12944 & 0.12944 & 0.12944 & 0.12944 & 0.12944 \\ 0.08185 & 0.08185 & 0.08185 & 0.08185 & 0.08185 & 0.08185 & 0.08185 & 0.08185 \\ 0.05540 & 0.05540 & 0.05540 & 0.05540 & 0.05540 & 0.05540 & 0.05540 & 0.05540 \\ 0.05000 & 0.05000 & 0.05000 & 0.05000 & 0.05000 & 0.05000 & 0.05000 & 0.05000 \\ 0.20104 & 0.20104 & 0.20104 & 0.20104 & 0.20104 & 0.20104 & 0.20104 & 0.20104 \\ 0.06855 & 0.06855 & 0.06855 & 0.06855 & 0.06855 & 0.06855 & 0.06855 & 0.06855 \end{bmatrix}$$

and

$$M^{32} = \begin{bmatrix} 0.18890 & 0.18890 & 0.18890 & 0.18890 & 0.18890 & 0.18890 & 0.18890 & 0.18890 \\ 0.22483 & 0.22483 & 0.22483 & 0.22483 & 0.22483 & 0.22483 & 0.22483 & 0.22483 \\ 0.12944 & 0.12944 & 0.12944 & 0.12944 & 0.12944 & 0.12944 & 0.12944 & 0.12944 \\ 0.08185 & 0.08185 & 0.08185 & 0.08185 & 0.08185 & 0.08185 & 0.08185 & 0.08185 \\ 0.05540 & 0.05540 & 0.05540 & 0.05540 & 0.05540 & 0.05540 & 0.05540 & 0.05540 \\ 0.05000 & 0.05000 & 0.05000 & 0.05000 & 0.05000 & 0.05000 & 0.05000 & 0.05000 \\ 0.20104 & 0.20104 & 0.20104 & 0.20104 & 0.20104 & 0.20104 & 0.20104 & 0.20104 \\ 0.06855 & 0.06855 & 0.06855 & 0.06855 & 0.06855 & 0.06855 & 0.06855 & 0.06855 \end{bmatrix}.$$

Notice that (i) accurate to five decimal places, $M^{16} = M^{32}$, and (ii) the columns of M^{16} are all the same! Higher powers of M will be no different.

This means that after 16 months, about 18.89% of bikes from station 31041 will end up at station 31041. 18.89% of bikes from station 31042 will end up at station 31041. 18.89% of bikes from station 31043 will end up at station 31041. In fact, about 18.89% of bikes from each station will end up at station 31041. Altogether, then, about 18.89% of all the bikes in the neighborhood will end up at station 31041. Likewise, about 22.48% of all the bikes in the neighborhood will end up at station 2, 12.94% at station 3, and so on. No matter how the bikes are initially distributed, they will end up distributed this way after some time, and stay that way (so long as the empirical migration percentages hold).

Crumpet 32: Why it Works

Suppose M is a positive column-stochastic matrix. The Gershgorin circle theorem ensures that M^T (and therefore M) has no eigenvalue with magnitude greater than one and its only possible eigenvalue with magnitude equal to one is 1. Letting $\mathbf{1}$ be the column vector whose entries are all 1, note that $M^T\mathbf{1} = \mathbf{1}$ (this is equivalent to saying the rows of M^T sum to 1) so 1 is indeed an eigenvalue of M^T. Hence M has dominant eigenvalue 1. Since $M^T - I$ is real, its null space admits a basis of real vectors. Suppose $\mathbf{w} \neq k\mathbf{1}$ is a real nonzero vector in the null space of $M^T - I$, and assume $\mathbf{w}$ has at least one positive entry (if it does not, multiply it by -1). Now set

$$\alpha_{max} = \max\{\alpha : \mathbf{1} - \alpha\mathbf{w} \text{ is nonnegative}\}$$

(the set is nonempty since it contains 0 and closed since the limit of a nonnegative sequence is nonnegative, so it has a maximum). Now $\mathbf{u} = \mathbf{1} - \alpha_{max}\mathbf{w}$ has at least one zero entry (if it does not, then α_{max} is not maximal). But $\mathbf{u}$ is then a nonnegative eigenvector (of the positive matrix M) and hence must be positive, contradicting that it has at least one zero entry. Thus no such $\mathbf{w}$ exists, and the eigenspace of 1 is one-dimensional. A dominant eigenvalue with a one-dimensional eigenspace is exactly what is needed for the power method to work (section 6.2). In this case, the matrix M is stochastic, so instead of computing $\mathbf{v}_k = M^k\mathbf{v}_0$, we simply calculate M^k. The entries of M^k will never tend to 0 or infintity since powers of stochastic matrices are stochastic, so we do not have to worry about scaling after each iteration. We can then multiply any nonzero $\mathbf{v}_0$ by M^k to find the approximation $\mathbf{v}_k$ of the dominant eigenvector.

Vocabulary

- A **positive matrix** is one whose entries are all positive.
- A **nonnegative matrix** is one who entries are all nonnegative.
- A **stochastic matrix** is a nonnegative matrix whose columns each sum to 1.

Lemmas

In each of the following lemmas, M is an $n \times n$ matrix.

- If M is positive and $\mathbf{w}$ is a nonnegative eigenvector, then $\mathbf{w}$ is positive.

  Proof. Since eigenvectors are nonzero and $\mathbf{w}$ is nonnegative, $\mathbf{w}$ has a positive entry, say its i^{th}. If follows that

  $$(M\mathbf{w})_{j,1} = \sum_{k=1}^{n} M_{j,k}\mathbf{w}_{k,1} \geq M_{j,i}\mathbf{w}_{i,1} > 0$$

  for each $j = 1, 2, \ldots, n$. □

- The eigenvalues of M and M^T are the same.

  Proof. Since the determinants of a matrix and its transpose are equal (section 3.5), for any scalar λ,

  $$\det(M - \lambda I) = \det(M - \lambda I)^T = \det(M^T - (\lambda I)^T) = \det(M^T - \lambda I).$$

  Hence M and M^T have the same characteristic equation and therefore the same eigenvalues. □

- If the entries of M are real numbers and λ is a real number, then the null space of $M - \lambda I$ admits a basis of vectors with real entries.

  Proof. Since the null space of any matrix can be found through row reduction, and row reducing a real matrix does not requre complex numbers, a real basis of $M - \lambda I$ (which has real entries) exists. □

- Gershgorin circle theorem: Every (possibly complex) eigenvalue of M lies in at least one disk with center $M_{i,i}$ and radius $r_i = \sum_{j \neq i} |M_{i,j}|$, $i = 1, 2, \ldots, n$.

  Proof. Suppose $\lambda, \mathbf{w}$ is an eigenpair of M and let $\mathbf{w}_{i,1}$ be the entry of $\mathbf{w}$ with greatest magnitude. Because $M\mathbf{w} = \lambda\mathbf{w}$, we have for each $i = 1, 2, \ldots, n$, $\lambda x_{i,1} = \sum_{j=1}^{n} M_{i,j} x_{j,1}$. Hence

  $$(\lambda - M_{i,i})x_{i,1} = \lambda x_{i,1} - M_{i,i}x_{i,1} = \sum_{j \neq i} M_{i,j}x_{j,1}$$

  from which it follows

  $$\left|\lambda - M_{i,i}\right| = \frac{1}{|x_{i,1}|}\left|\sum_{j \neq i} M_{i,j}x_{j,1}\right| \leq \sum_{j \neq i} |M_{i,j}| \cdot \left|\frac{x_{j,1}}{x_{i,1}}\right| \leq \sum_{j \neq i} |M_{i,j}|.$$

  □

- Powers of stochastic matrices are stochastic.

  Proof. Let S be a stochastic matrix. By definition, $\mathbf{1}^T S = \mathbf{1}^T$ (the sum of each column of S is one). By induction, assume S^k is stochastic for some $k \geq 1$. Then

  $$\mathbf{1}^T S^{k+1} = \left(\mathbf{1}^T S^k\right) S = \mathbf{1}^T S = \mathbf{1}^T$$

  so the columns of S^{k+1} sum to one. Of course powers of nonnegative matrices are nonnegative, so S^{k+1} is stochastic. □

For example, suppose Capital Bikeshare supplies each station with the same number of bicycles to begin. That information can be recorded in a column vector with length eight and each entry equal to $\frac{1}{8}$. After one month, we assume that the empirical data on transitions from one station to another is reasonably accurate, so the distribution of bicycles will be approximately

$$M\begin{bmatrix}\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\end{bmatrix}\approx\begin{bmatrix}0.166\\0.226\\0.116\\0.103\\0.071\\0.062\\0.185\\0.071\end{bmatrix}.$$

Multiplying the distribution vector by the **transition matrix** M gives the new distribution of bicycles among the stations. After another month, the distribution of bicycles will be approximately

$$M\begin{bmatrix}0.166\\0.226\\0.116\\0.103\\0.071\\0.062\\0.185\\0.071\end{bmatrix}\approx M^2\begin{bmatrix}\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\end{bmatrix}\approx\begin{bmatrix}0.180\\0.226\\0.128\\0.086\\0.058\\0.052\\0.196\\0.073\end{bmatrix}$$

and after four months,

$$M^2\begin{bmatrix}0.180\\0.226\\0.128\\0.086\\0.058\\0.052\\0.196\\0.073\end{bmatrix}\approx M^4\begin{bmatrix}\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\end{bmatrix}\approx\begin{bmatrix}0.187\\0.225\\0.129\\0.082\\0.056\\0.050\\0.201\\0.069\end{bmatrix}$$

and after eight months,

$$M^4\begin{bmatrix}0.187\\0.225\\0.129\\0.082\\0.056\\0.050\\0.201\\0.069\end{bmatrix}\approx M^8\begin{bmatrix}\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\\\frac{1}{8}\end{bmatrix}\approx\begin{bmatrix}0.189\\0.225\\0.129\\0.082\\0.055\\0.050\\0.201\\0.069\end{bmatrix}$$

which is, accurate to three decimal places, equal to the columns of M^{16} (and M^{32} for that matter). The distribution of bicycles does not change much after the first two months.

More importantly, the distribution we see after 8 months will be the eventual distribution of bicycles no matter the initial distribution! Given any initial distribution of bicycles,

$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 & w_5 & w_6 & w_7 & w_8 \end{bmatrix}^T$$

(where $\sum_{i=1}^{8} w_i = 1$ and each w_i is nonnegative),

$$M^{16}\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ w_8 \end{bmatrix} \approx \begin{bmatrix} 0.18890 \sum_{i=1}^{8} w_i \\ 0.22483 \sum_{i=1}^{8} w_i \\ 0.12944 \sum_{i=1}^{8} w_i \\ 0.08185 \sum_{i=1}^{8} w_i \\ 0.05540 \sum_{i=1}^{8} w_i \\ 0.05000 \sum_{i=1}^{8} w_i \\ 0.20104 \sum_{i=1}^{8} w_i \\ 0.06855 \sum_{i=1}^{8} w_i \end{bmatrix} = \sum_{i=1}^{8} w_i \begin{bmatrix} 0.18890 \\ 0.22483 \\ 0.12944 \\ 0.08185 \\ 0.05540 \\ 0.05000 \\ 0.20104 \\ 0.06855 \end{bmatrix} = \begin{bmatrix} 0.18890 \\ 0.22483 \\ 0.12944 \\ 0.08185 \\ 0.05540 \\ 0.05000 \\ 0.20104 \\ 0.06855 \end{bmatrix}.$$

This means the distribution of bicycles in the long-run is given by this vector, $\mathbf{v} = M_{i,1}$, the eigenvector of M corresponding to eigenvalue 1. This distribution is called the steady-state distribution because if reached, it never deviates: $M\mathbf{v} = \mathbf{v}$, so $\mathbf{v}$ remains steady over time. We would expect to see the bicycles distributed among the stations in these proportions after adequate time.

Formalities

The sequence of bicycle distributions in the Capital Bikeshare scenario is an example of a discrete-time **Markov chain**. Any Markov chain necessarily consists of a set of **states** (stations in our example), a set of probabilities that some object (bicycle in our example) will transition from one state to any other depending only on its current state after one time step, and an initial distribution among the possible states. The matrix M, where $M_{i,j}$ is the probability of transitioning from state j to state i, is the **transition matrix**, and each distribution $\mathbf{v}_n$ other than $\mathbf{v}_0$ satisfies $\mathbf{v}_n = M\mathbf{v}_{n-1}$.

Other situations that can be modeled by Markov chains include

1. board games whose movement is determined by the roll of a die, such as Snakes and Ladders (the spaces on the board are the states, the game piece is the object, and the roll of the die provides the transition probabilities);
2. sentence construction, as used by computer auto-completion (the states are the words of a specific language, the object is the reader's focus, and the probability of one word following another in a sentence provides the transition probabilites);
3. a closed economy—one where a set of commodities is produced and consumed by the same group (the states are the sectors of the economy, the commodities are the objects, and transitioning is interpreted as consumption);
4. the weather (the states are weather conditions such as sunny, cloudy, and rainy, the object is the weather, and the transition probabilities are the conditional probabilities that one weather condition will follow another on, say, the next day);
5. gambling (the states are the amounts of money the gambler could have, the object is the gambler, and the likelihoods of winning or losing certain amounts of money provide the transition probabilities);
6. arrival or service times in a single server queue (the states are the possible sizes of the queue, the object is the queue, and the likelihoods of increasing or decreasing the size of the queue by certain amounts provide the transition probabilities).

Key Concepts

transition matrix a square matrix M where $M_{i,j}$ is the probability of transitioning from state j to state i in one time step. The entries of M are nonnegative and each column of M sums to 1.[2]

Markov chain a sequence of distributions arising from an initial distribution $\mathbf{v}_0$ and the recurrence $\mathbf{v}_n = M\mathbf{v}_{n-1}$, $n > 0$ for some transition matrix M.[3]

[2] A square matrix whose entries are nonnegative and whose columns each sum to 1 is also called a **stochastic matrix** (whether it models state transition probabilities or not).

[3] In a more general setting, the transition matrix may change with time, and would then be replaced by M_n.

state one of the possible conditions of the object associated with a Markov chain.

steady-state distribution a distribution $\mathbf{v}$ such that $M\mathbf{v} = \mathbf{v}$ (by definition an eigenvector corresponding to eigenvalue 1).

properties of a transition matrix If M is a transition matrix,

1. each column of M sums to one;
2. each entry of M is nonnegative;
3. 1 is one of its eigenvalues;
4. none of its eigenvalues has magnitude greater than one.

If, additionally, all the entries of some power, M^k, of M are all positive,

1. 1 is a dominant eigenvalue;
2. the eigenspace of 1 is one-dimensional;
3. each column of M^k approaches the same eigenvector, that corresponding with the eigenvalue 1, the steady-state vector.

Exercises

1. Is M a transition matrix?

   (a) $M = \begin{bmatrix} 0.492 & 0.118 & 0.516 \\ 0.346 & 0.819 & 0.361 \\ 0.472 & 0.063 & 0.123 \end{bmatrix}$

   (b) $M = \begin{bmatrix} 0.09 & 0.686 & 0.168 \\ 0.908 & 0.036 & 0.807 \\ 0.002 & 0.278 & 0.485 \end{bmatrix}$ [S]-339

   (c) $M = \begin{bmatrix} 0.409 & 0.696 \\ 0.179 & 0.156 \\ 0.412 & 0.148 \end{bmatrix}$

   (d) $M = \begin{bmatrix} 0.826 & 0.006 & 0.235 \\ 0.104 & 0.122 & 0.609 \\ 0.07 & -0.72 & 0.156 \end{bmatrix}$ [A]-362

   (e) $M = \begin{bmatrix} 0.485 & 0.145 & -0.58 \\ 0.302 & 0.46 & -0.22 \\ 0.213 & 0.395 & -0.20 \end{bmatrix}$

   (f) $M = \begin{bmatrix} 0.341 & 0.104 & 0.217 \\ 0.525 & 0.592 & 0.249 \\ 0.134 & 0.304 & 0.534 \end{bmatrix}$ [A]-362

2. Given transition matrix M and initial distribution vector $\mathbf{v}_0$, (i) calculate M^2 and M^3; (ii) determine the probability of transitioning from state 1 to state 2 over the course of one time step, two time steps, and three time steps; (iii) determine the probability of being in state 2 after three time steps given equal likelihood of starting in any state; (iv) calculate $\mathbf{v}_3$; and (v) for the given initial distribution, $\mathbf{v}_0$, what is the probability of being in state 2 after three time steps?

   (a) SageMathCell 117 $M = \begin{bmatrix} 0.983 & 0.139 \\ 0.017 & 0.861 \end{bmatrix}$; $\mathbf{v}_0 = \begin{bmatrix} 0.484 \\ 0.516 \end{bmatrix}$

   (b) SageMathCell 118 $M = \begin{bmatrix} 0.053 & 0.846 \\ 0.947 & 0.154 \end{bmatrix}$; $\mathbf{v}_0 = \begin{bmatrix} 0.653 \\ 0.347 \end{bmatrix}$ [A]-362

   (c) SageMathCell 119 $M = \begin{bmatrix} 0.484 & 0.653 \\ 0.516 & 0.347 \end{bmatrix}$; $\mathbf{v}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

   (d) SageMathCell 120 $M = \begin{bmatrix} 0.403 & 0.357 & 0.024 \\ 0.446 & 0.116 & 0.24 \\ 0.151 & 0.527 & 0.736 \end{bmatrix}$; $\mathbf{v}_0 = \begin{bmatrix} 0.105 \\ 0.019 \\ 0.876 \end{bmatrix}$

   (e) SageMathCell 121 $M = \begin{bmatrix} 0.151 & 0.355 & 0.163 \\ 0.21 & 0.637 & 0.56 \\ 0.639 & 0.008 & 0.277 \end{bmatrix}$; $\mathbf{v}_0 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ [A]-362

   (f) SageMathCell 122 $M = \begin{bmatrix} 0.385 & 0.3 & 0.429 \\ 0.049 & 0.651 & 0.378 \\ 0.566 & 0.049 & 0.193 \end{bmatrix}$; $\mathbf{v}_0 = \begin{bmatrix} 0.578 \\ 0.269 \\ 0.153 \end{bmatrix}$

3. For any square matrix M, explain why M and M^T have the same eigenvalues. [A]-362

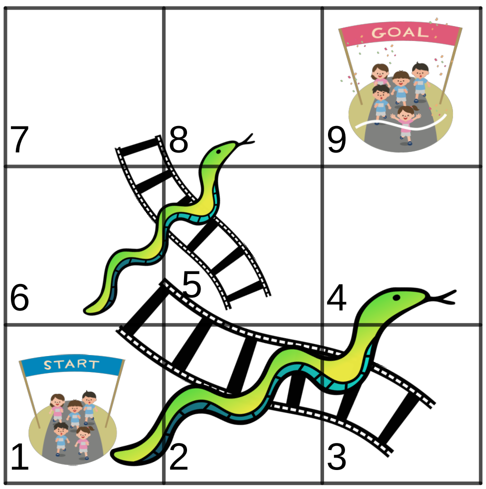
4. Every transition matrix is guaranteed to have 1 as an eigenvalue (M^T has 1 as an eigenvalue since $M^T\mathbf{1} = \mathbf{1}$).

Find the eigenspace of 1 for the transition matrix in question 2. [$\mathbb{A}$]-362

5. Calculate M^{32} for the transition matrix of question 2 and compare it to your answer in question 4. [$\mathbb{A}$]-363

6. Find the steady-state distribution for the transition matrix in question 2. [$\mathbb{A}$]-363

7. For the simplified Snakes and Ladders board[4], a 6-sided die with only the numbers 1,2, and 3 on it is used (two of each, making rolling a 1,2, or 3 equally likely).



The rules for Snakes and Ladders can be found at GameRules.com. [$\mathbb{A}$]-363

(a) Create a 5×5 transition matrix. Each block on the board represents one state (location of the playing piece). It is impossible to end a turn on spaces 3,4,5, or 8 since the playing piece immediately slides up or down from there, so only the other 5 states need to be included. Transitioning from state 9 is represented by a column with 4 zeros and a one, indicating that once square 9 is reached the playing piece never leaves! Mathematically, state 9 is called an **absorbing state** (a state that once entered cannot be left).

(b) SageMathCell What is the likelihood of reaching the goal (square 9) in one roll? two rolls? three rolls? ten rolls?

(c) Assuming the columns (and rows) of the transition matrix represent states 1,2,6,7,9 in that order, the eigenvector corresponding to the eigenvalue 1 is $\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T$. What does this mean in the context of the game?

8. Create your own Snakes and Ladders board and repeat the analysis of question 7.

9. The Leontief model for a closed economy is similar to a Markov chain where each column of the transition matrix represents a sector's consumption of each commodity as a proportion of the economy. Consequently the matrix is commonly called a consumption matrix rather than a transition matrix. Suppose an economy has three industries: farming, building, and clothing. For every dollar of food produced, the farmers use \$0.375, the builders use \$0.35, and the tailors use \$0.275. For every dollar of building, the builders use \$0.286, the farmers use \$0.214, and the tailors use \$0.5. For every dollar of clothing produced, the tailors use \$0.348, the builders use \$0.326, and the farmers use \$0.326. [$\mathbb{A}$]-363

(a) Since all of the production of these three industries is consumed by the three industries themselves, the consumption matrix has columns that sum to 1. Write down the consumption matrix for this economy.

(b) A production vector $\mathbf{v}$ represents the production of each industry in dollars. What does the vector $M\mathbf{v}$ represent?

(c) Let $\mathbf{v} = \begin{bmatrix} 10 & 57 & 33 \end{bmatrix}^T$ and calculate $M\mathbf{v}$. At these production levels (which are in thousands of dollars), is there a sector that consumes more than it produces? How about a sector that produces more than it consumes?

(d) How do you know there is an "everybody is happy" production vector $\mathbf{v}$, one such that $M\mathbf{v} = \mathbf{v}$?

(e) Calculate the "everybody is happy" production vector. In terms of the economy, why might it be referred to as such?

(f) Any solution of $M\mathbf{v} = \mathbf{v}$ is, in the context of the problem, called an equilibrium. Explain in terms of the economy what is in equilibrium at production levels in such a $\mathbf{v}$?

10. Quito, the capital of Ecuador, is located essentially on the equator, and therefore does not experience seasons. The weather is more or less the same year round—average low temperature about 50°F and average high temperature about 68°F all 12 months of the year. Clouds and rain are similarly predictable. Suppose you are visiting Quito and a friend of yours claims it rains there one quarter of the time. For two months you record the weather. Your data suggest that on a rainy day there is a $\frac{3}{14}$ chance it will be rainy again the next day and on a dry day there is a $\frac{5}{7}$ chance it will be dry again the next day. Are your observations consistent with your friend's claim? Create a Markov chain model of the weather and use it to answer the question.

11. In a very simple gambling game, you win one dollar with probability $\frac{2}{5}$ and lose one dollar with probability $\frac{3}{5}$. You start with one dollar, and the game ends when you either go broke or win 3 dollars (have 4 dollars).

(a) Model the playing of this game with a Markov chain. The states are the numbers of dollars you can have and the transition matrix is formed from the probabilities of going from one amount of money to another. What is the transition matrix, M?

[4] Snake, ladder, start, and goal clipart from PublicDomainVectors.org.

(b) What is the probability the game will last more than 10 rounds? That is, with what probability will you have neither \$0 nor \$4 after 10 rounds?

(c) After 10 rounds, what is the probability you have gone broke?

(d) After 10 rounds, what is the probability you have won \$3?

(e) Should you play this game?

(f) Find basis vectors for the eigenspace of eigenvalue 1.

(g) Compare the basis vectors to the columns of M^{64}. What do you notice?

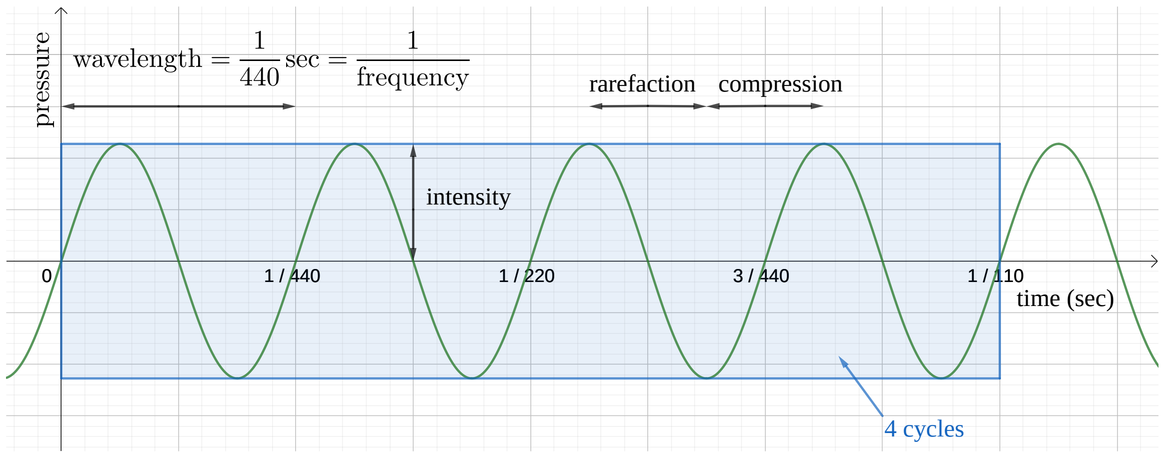
(h) Let p be the probability of winning one dollar in a round. What value of p gives you a 50% chance of winning \$3 within 10 rounds?

(i) How does p change if the game ends only when you either go broke or win \$4 instead of \$3? Is the value of p that gives you a 50% chance of winning \$4 within 10 rounds greater or less than before?

12. Redo the bikeshare analysis for 2020 in the same section of Alexandria as done in the text. Data can be retrieved at the ancillary website. How different is the expected bike distribution in 2020? Can you think of a reason the data are likely to be very different from the 2018 data?

13. Redo the bikeshare analysis for 2018 in a different neighborhood. Data can be retrieved at the ancillary website.

Figure 7.3.1: Mathematical Model of a 440 Hz Simple Harmonic Sound Wave



7.3 Fourier Series [4.6, 6.4, calculus]

Sound is the perception of pressure variation. Tuning forks, speakers, musical instruments, voice boxes, whistles, and anything else that makes sound must somehow cause varying pressure. One common way to create pressure variation is through physical vibration. Tuning forks, speakers, the strings of stringed instruments, and vocal cords all use this technique. Their vibrations cause alternating moments of **compression** (increased pressure) and **rarefaction** (reduced pressure) in the air.

The greater the difference between high and low pressures, the louder the sound. Sometimes the pressure difference, called **volume** or **intensity**, is so great, our whole bodies vibrate. Thunder, subwoofers, fireworks, and helicopters can do this, but for most sounds it is only our eardrums that perceive the pressure variation.

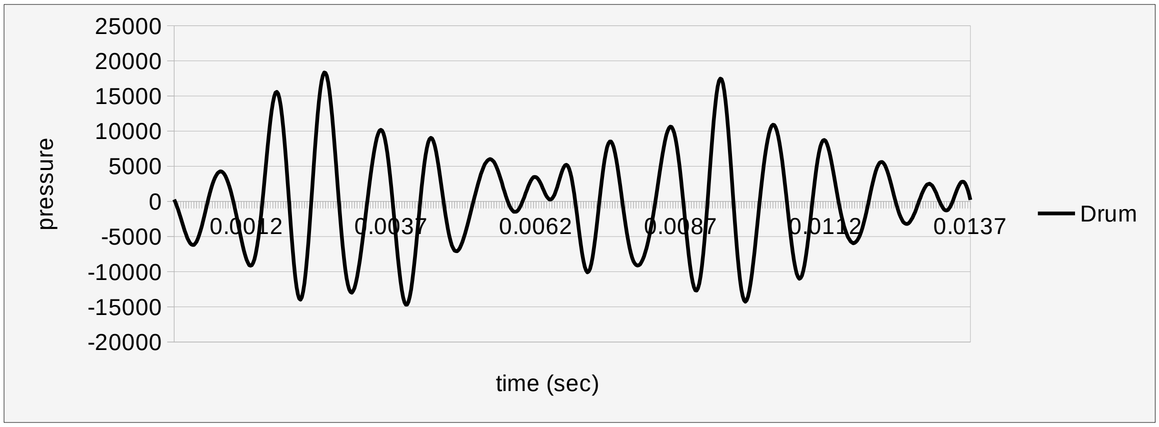
The faster the pressure alternates between high and low, the higher the **pitch**. Middle C, for example, is the result of pressure varying from neutral to high to low and back to neutral approximately 261.6 times per second. Each variation through neutral, high, low and back to neutral is one **cycle**, so we also say that middle C has a frequency of 261.6 cycles per second. One Hertz, abreviated Hz, equals one cycle per second, so we also say middle C has a frequency of 261.6 Hz. The highest note on a piano, B_8, has a frequency of about 7902.1 Hz and the lowest note on a piano has a frequency of about 16.35 Hz.

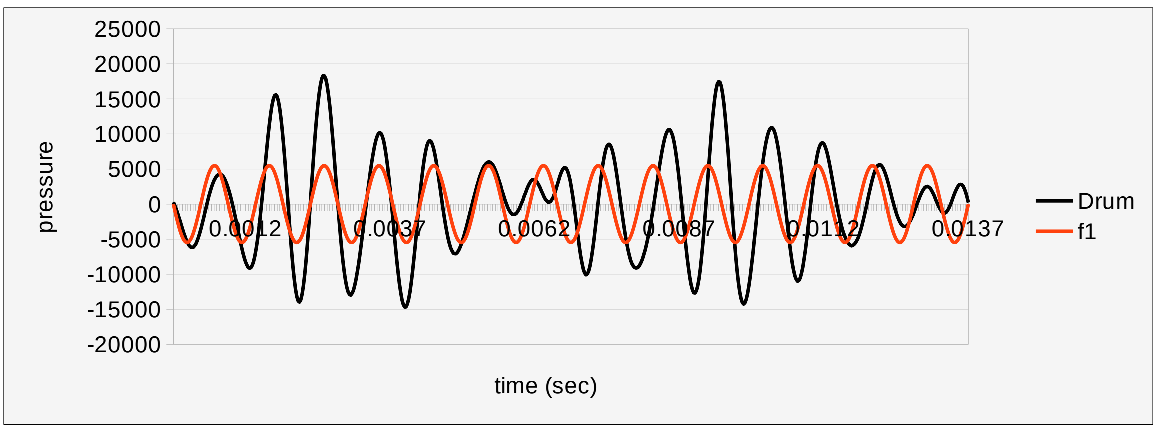
The human ear is capable of perceiving frequencies between about 20 Hz and 20,000 Hz. Air pressure can certainly alternate between high and low pressures slower than 20 cycles per second and faster than 20,000 cycles per second. Our ears just won't pick up those vibrations. Dog whistles emit sound between 23,000 Hz and 54,000 Hz[5], all above the range of human hearing but within the range of canine hearing. Elephants were the first land animals to be observed to produce sound below the range of human hearing[19], creating calls with frequencies as low as 14 Hz.

A sound wave can be modeled by a record of the pressure it causes on a receiver such as a microphone or eardrum over time. The simplest type of sound waves are **simple harmonic** vibrations–one intensity, one frequency, shifting from high to low pressure smoothly as a sine curve. Until the advent of electronics, the closest approximation to a simple harmonic vibration was the sound of a tuning fork. Over the course of a second or so, neither the frequency nor the intensity of a vibrating tuning fork changes appreciably, and the vibrations are sinusoidal. The graph of a 440 Hz sine wave is a mathematical model of this sound. See figure 7.3.1.

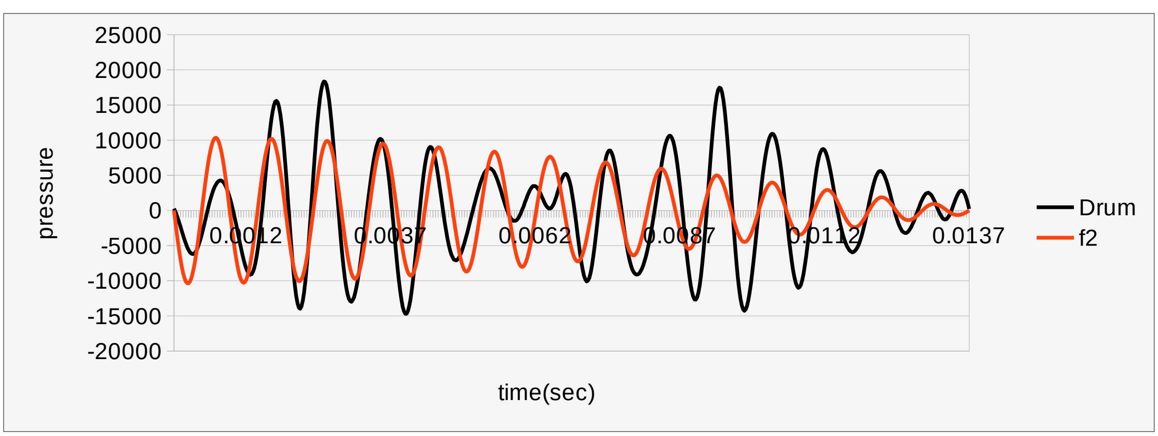
Naturally produced sounds are not so neat. Even a tuning fork does not produce sound in a perfect sine wave. Its intensity decreases continuously as it rings, and air particles do not compress and rarefy in a perfectly sinusoidal pattern. The matter is even more complex for musical instruments. Even on a stringed instrument where a string vibrates at a steady frequency, the body picks up the vibration and imparts its own signature frequencies to the sound. Wind and percussion instruments are the same. The richness of their sound comes from a variable intensity patchwork of many frequencies. The graph of two cycles of an actual recording of a singing drum are shown in figure 7.3.2. The wave is clearly not sinusoidal, featuring 8 peaks and 8 valleys per cycle. In a sense, the best sinusoidal approximation of this sound wave is shown below as $f1(t) = -5498.21 \sin\left(29 \cdot \frac{8820}{121}\pi t\right)$, a frequency of $\frac{1}{2} \cdot 29 \cdot \frac{8820}{121} \approx 1056.9$ Hz.

Figure 7.3.2: Two cycles of the sound of a drum.





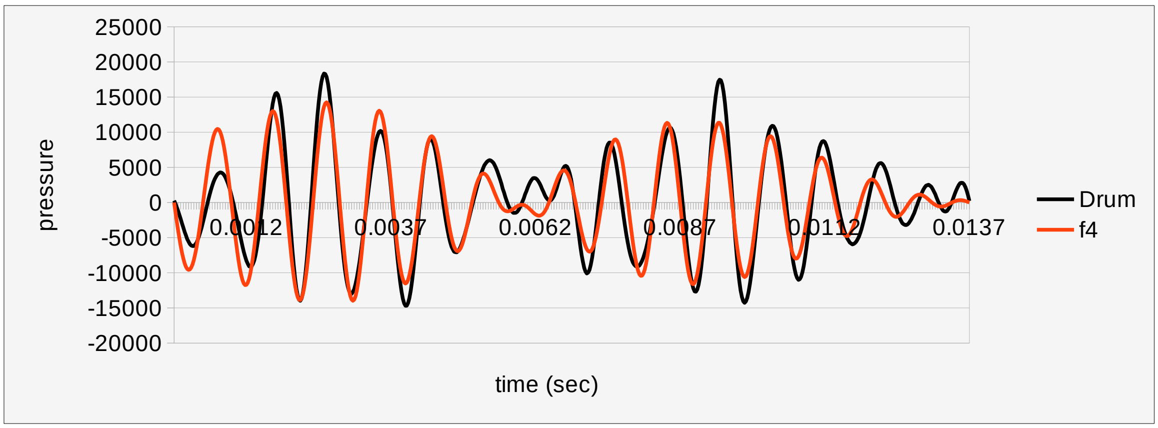
As expected, this sine wave does a poor job of approximating the drum wave. But we can do better. Allowing the sum of two sine waves, we can improve the approximation considerably, shown below as $f2(t) = -5498.21 \sin\left(29 \cdot \frac{8820}{121}\pi t\right) - 4891.4 \sin\left(28 \cdot \frac{8820}{121}\pi t\right)$, a combination of frequencies 1056.9 Hz and $\frac{1}{2} \cdot 28 \cdot \frac{8820}{121} \approx 1020.5$ Hz.



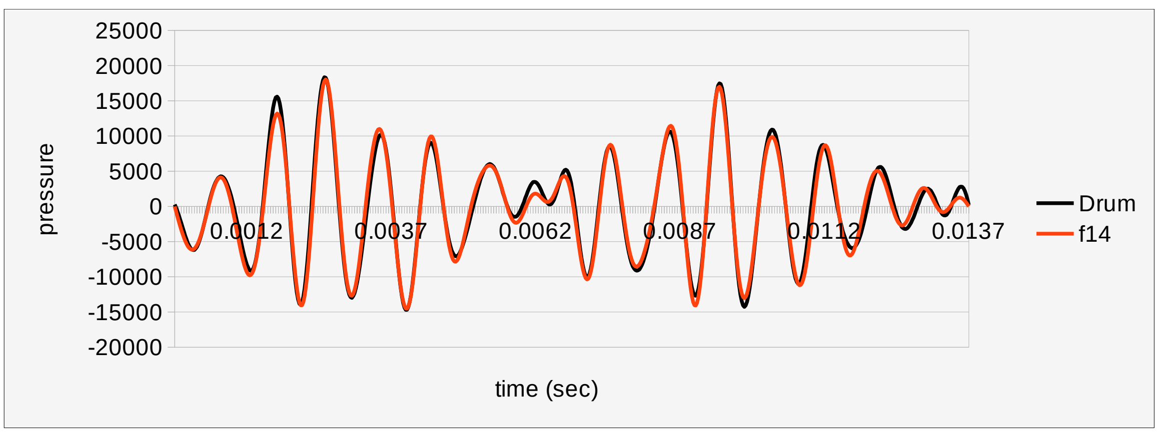
The approximation now peters out as does the drum wave, and the peaks and valleys match better. Allowing a combination of four sinusoidal waves,

$$f4(t) = -5498.21 \sin\left(29 \cdot \frac{8820}{121}\pi t\right) - 4891.4 \sin\left(28 \cdot \frac{8820}{121}\pi t\right)$$
$$+ 4394.8 \sin\left(33 \cdot \frac{8820}{121}\pi t\right) - 3469.0 \sin\left(31 \cdot \frac{8820}{121}\pi t\right)$$

the approximation continues to improve, as seen here.



The more sinusoidal waves allowed, the better the approximation. The differences largely disappear with the allowance of 14 sinusoidal waves:



In order of greatest to least intensity, the frequencies represented in $f14$ are 1056.9, 1020.5, 1202.7, 1129.8, 984.0, 1166.3, 291.6, 1640.1, 1093.4, 1457.9, 1567.2, 1275.6, 328.0, and 1239.2 Hz. For the brief moment represented in the graph $\left(\frac{121}{8820} \approx 0.01372 \text{ sec}\right)$, this means the sound of the drum was dominated by these 14 frequencies.

Similarly examining a full 2 seconds of the audio reveals that the overall dominant frequencies of this particular drum sound, in order from most to least dominant are 290, 1177, and 1027 Hz. The note played was likely D_4, whose frequency is 293.66 Hz, equivalent to the D string on a violin or viola played open.

But how do we know which frequencies and intensities to use in approximation? In principle, the answer is simple. Theorem 18 of section 6.4 provides the guidance. Orthogonal projection of the function gives the right intensities. All we need are an appropriate inner product space and a basis for some subspace. Except in extreme cases, pressure waves, and therefore sound waves, vary continuously, so it makes sense to consider the vector space of continuous functions over some interval. In particular, we will consider subspaces of $C([0, L])$, the set of all functions which are continuous on the closed interval $[0, L]$.

Much like the inner products of exercises 1e and 1f of section 4.6, $\langle f, g \rangle = \int_0^L f(x)g(x)\,dx$ defines an inner product on $C([0, L])$. Can you justify it? Answer on page 255.

Vectors in $C([0, L])$ are continuous functions, so a basis for any subspace will have to be a collection of functions that are continuous on $[0, L]$. As suggested by their use earlier, collections of trigonometric functions provide convenient bases. For all $m = 0, 1, 2, \ldots$ and all $n = 1, 2, \ldots$

$$\left\langle \cos\left(\frac{m\pi}{L}t\right), \cos\left(\frac{n\pi}{L}t\right)\right\rangle = 0 \tag{7.3.1}$$

and

$$\left\langle \sin\left(\frac{m\pi}{L}t\right), \sin\left(\frac{n\pi}{L}t\right)\right\rangle = 0 \tag{7.3.2}$$

whenever $m \neq n$. That is, for any distinct positive integers m and n, $\cos\left(\frac{m\pi}{L}t\right)$ and $\cos\left(\frac{n\pi}{L}t\right)$ are orthogonal, as are $\sin\left(\frac{m\pi}{L}t\right)$ and $\sin\left(\frac{n\pi}{L}t\right)$. Inner products (7.3.1) and (7.3.2) can be verified with the help of two trigonometric identities. Recall that $\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$, so

$$\frac{1}{2}\left[\cos(\alpha - \beta) + \cos(\alpha + \beta)\right] = \cos\alpha\cos\beta \tag{7.3.3}$$

and

$$\frac{1}{2}\left[\cos(\alpha - \beta) - \cos(\alpha + \beta)\right] = \sin\alpha\sin\beta. \tag{7.3.4}$$

Can you use (7.3.4) to show (7.3.2)? Answer on page 255. And now we have our candidates for a vector space, $C([0, L])$, and basis elements, $\sin\left(\frac{m\pi}{L}t\right)$ with $m > 0$ or $\cos\left(\frac{m\pi}{L}t\right)$ with $m \geq 0$.

When the function of interest takes the value zero at both endpoints of the interval, as is the case for the sound waves we have looked at, it makes best sense to use sine functions for a basis. Every sine function of the form $\sin\left(\frac{m\pi}{L}t\right)$ takes the value zero at both $t = 0$ and $t = L$ so the approximation is exact at the endpoints no matter how many basis elements are used. If the function were nonzero at either endpoint, it would make more sense to use cosine functions for a basis as this allows approximation of the nonzero endpoint(s).

Crumpet 33: A theorem of Fejér

[Fejér, 1900] *Let f be a continuous function on $[-\pi, \pi]$ for which $f(-\pi) = f(\pi)$. Then the sequence of Cesàro means of the partial sums of the Fourier series for f converges uniformly to f on* $[-\pi, \pi]$.[22] This theorem applies to any continuous function on $[0, \pi]$ by extending it as an even function over $[-\pi, \pi]$ (in which case the sine terms all have zero Fourier coefficients, yielding a cosine series) or, when $f(0) = f(\pi) = 0$, as an odd function over $[-\pi, \pi]$ (in which case the cosine terms all have zero Fourier coefficients, yielding a sine series). If f additionally has a piecewise continuous first derivative, then the sequence of partial sums of the Fourier series for f converge uniformly to f on $[-\pi, \pi]$. For many physical applications, such as the sound waves discussed here, the functions we are trying to approximate have continuous first derivatives and therefore their extensions have piecewise continuous first derivatives, and the theorem applies. By proper scaling, these results can be modified to apply to domains such as $[-L, L]$ or $[0, L]$.

The upshot of results like this is that there exist bases of trigonometric functions for subspaces containing vectors (linear combinations of trigonometric functions) arbitrarily close to a given vector (function). In other words, certain functions can be approximated with arbitrary precision using sums of sines and cosines.

Given a continuous function, we choose a set of sine functions or a set of cosine functions as basis for a subspace and we project onto the subspace to find the best approximation. For example, consider approximating $f(x) = x(x-1)(x-2)$ over the interval $[0, 2]$. How closely can we approximate f by vectors in the spans of

1. $\mathcal{B}_1 = \left\{1, \cos\left(\frac{\pi}{2}t\right), \cos(\pi t)\right\}$ (from the family of cosine functions with $m = 0, 1, 2$)?

   Answer: The best approximation is $\mathbf{v} = \operatorname{proj}_{\operatorname{span}\{\mathcal{B}_1\}} f$:

$$\frac{\langle f, 1\rangle}{\langle 1, 1\rangle}1 + \frac{\left\langle f, \cos\left(\frac{\pi}{2}t\right)\right\rangle}{\left\langle \cos\left(\frac{\pi}{2}t\right), \cos\left(\frac{\pi}{2}t\right)\right\rangle}\cos\left(\frac{\pi}{2}t\right) + \frac{\langle f, \cos(\pi t)\rangle}{\langle \cos(\pi t), \cos(\pi t)\rangle}\cos(\pi t).$$

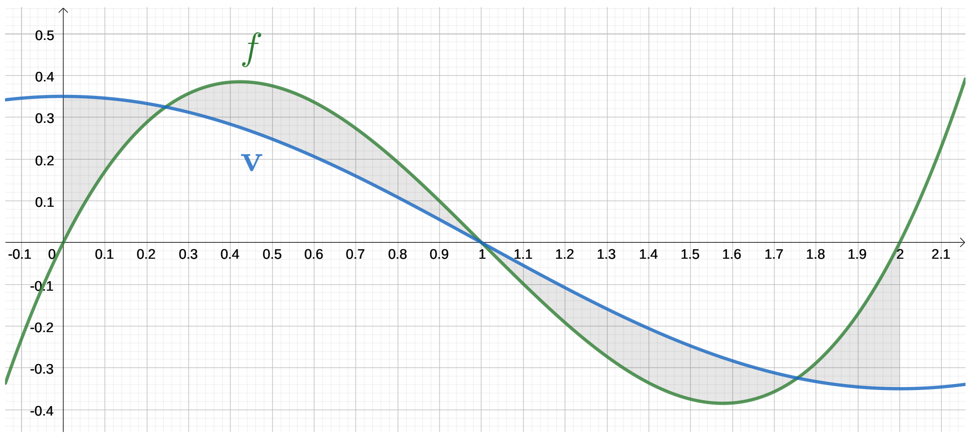
Using a computer algebra system to help with the integration,

$$\mathbf{v} = -16\frac{\pi^2 - 12}{\pi^4}\cos\left(\frac{\pi}{2}t\right).$$

Due to symmetry, the first and third terms are zero. The distance between $\mathbf{v}$ and f (see section 4.6) is

$$d(f, \mathbf{v}) = \|f - \mathbf{v}\| = \sqrt{\langle f - \mathbf{v}, f - \mathbf{v}\rangle} \approx 0.0299.$$

The graphs of f and $\mathbf{v}$ are shown below with the area between the two shaded over the interval $[0, 2]$. Since the distance between f and $\mathbf{v}$ involves integrating $(f - \mathbf{v})^2$, the shaded area does not represent the distance between the vectors exactly, but it helps give a visual sense of this distance.



2. $\mathcal{B}_2 = \left\{\sin\left(\frac{\pi}{2}t\right), \sin(\pi t), \sin\left(\frac{3\pi}{2}t\right)\right\}$ (from the family of sine functions with $m = 1, 2, 3$)?

Answer: The best approximation is $\mathbf{w} = \text{proj}_{\text{span}\{\mathcal{B}_2\}} f$:

$$\frac{\left\langle f, \sin\left(\frac{\pi}{2}t\right)\right\rangle}{\left\langle \sin\left(\frac{\pi}{2}t\right), \sin\left(\frac{\pi}{2}t\right)\right\rangle}\sin\left(\frac{\pi}{2}t\right) + \frac{\langle f, \sin(\pi t)\rangle}{\langle \sin(\pi t), \sin(\pi t)\rangle}\sin(\pi t) + \frac{\left\langle f, \sin\left(\frac{3\pi}{2}t\right)\right\rangle}{\left\langle \sin\left(\frac{3\pi}{2}t\right), \sin\left(\frac{3\pi}{2}t\right)\right\rangle}\sin\left(\frac{3\pi}{2}t\right).$$

Using a computer algebra system to help with the integration,

$$\mathbf{w} = \frac{12}{\pi^3}\sin(\pi x).$$

Due to symmetry the first and third terms are zero again. The distance between $\mathbf{w}$ and f is

$$d(f, \mathbf{w}) = \sqrt{\langle f - \mathbf{w}, f - \mathbf{w}\rangle} \approx 0.0026.$$

The graphs of f and $\mathbf{w}$ are shown below with the area between the two shaded. As the distance calculation suggests, this approximation is closer than the last.

3. $\mathcal{B}_3 = \left\{\sin\left(\frac{m\pi}{2}t\right) : m = 1, 2, \ldots, 10\right\}$?

Answer: The best approximation is $\mathbf{x} = \mathrm{proj}_{\mathrm{span}\{\mathcal{B}_3\}} f$:

$$\sum_{m=1}^{10} \frac{\left\langle f, \sin\left(\frac{m\pi}{2}t\right)\right\rangle}{\left\langle \sin\left(\frac{m\pi}{2}t\right), \sin\left(\frac{m\pi}{2}t\right)\right\rangle} \sin\left(\frac{m\pi}{2}t\right).$$

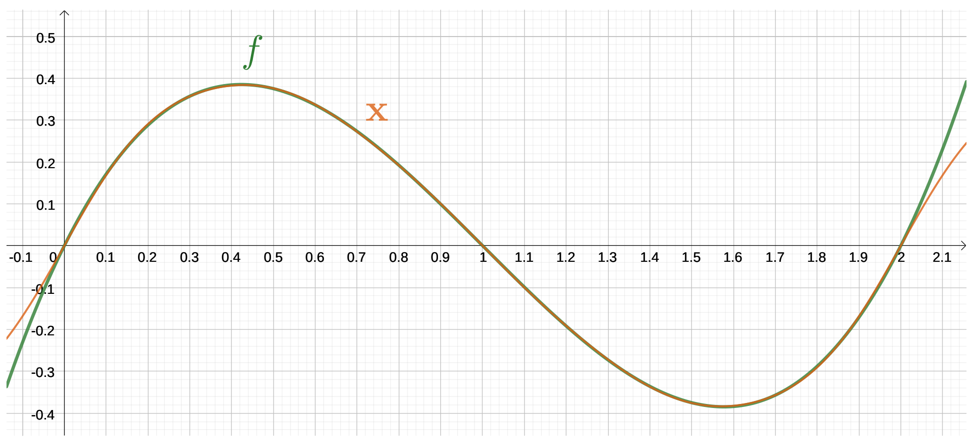
Using a computer algebra system to help with the integration,

$$\mathbf{x} = \frac{12}{125\pi^3}\sin(5\pi x) + \frac{3}{16\pi^3}\sin(4\pi x) + \frac{4}{9\pi^3}\sin(3\pi x) + \frac{3}{2\pi^3}\sin(2\pi x) + \frac{12}{\pi^3}\sin(\pi x).$$

Due to symmetry the odd terms are zero. The distance between $\mathbf{x}$ and f is

$$d(f, \mathbf{x}) \approx 0.00000572.$$

The graphs of f and $\mathbf{x}$ are shown below. The area between the vectors over the interval $[0, 2]$ is essentially imperceptible.



The choice of subspace, and therefore its basis, makes all the difference in how close the given function might be approximated.

In the case of the sound wave that opened this section, coefficients

$$\frac{\left\langle \mathrm{Drum}, \sin\left(m\cdot\frac{8820}{121}\pi t\right)\right\rangle}{\left\langle \sin\left(m\cdot\frac{8820}{121}\pi t\right), \sin\left(m\cdot\frac{8820}{121}\pi t\right)\right\rangle}, \quad m = 1, 2, \ldots, 120$$

were calculated and sorted from greatest magnitude to least. In decreasing order, the top 14 magnitudes were from the coefficients corresponding to $m = 29, 28, 33, 31, 27, 32, 8, 45, 30, 40, 43, 35, 9, 34$, so the set

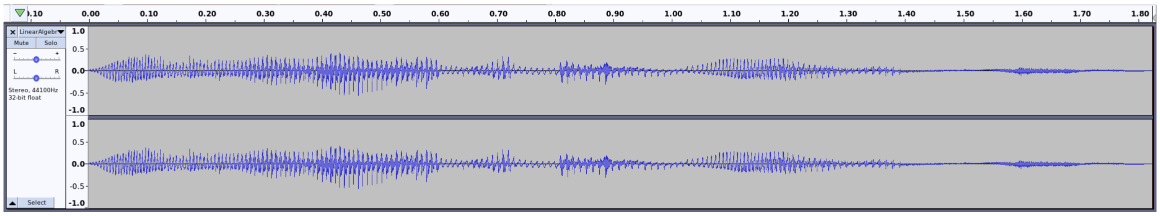
$$\left\{\sin\left(m\cdot\frac{8820}{121}\pi t\right) : m = 29, 28, 33, 31, 27, 32, 8, 45, 30, 40, 43, 35, 9, 34\right\}$$

and subsets were chosen as bases to create the approximations. The frequency of any harmonic function, $\sin(\omega x)$ or $\cos(\omega x)$, is $\frac{\omega}{2\pi}$, so the frequencies of these basis elements are

$$\begin{aligned}&\left\{m\cdot\frac{44100}{121} : m = 29, 28, 33, 31, 27, 32, 8, 45, 30, 40, 43, 35, 9, 34\right\}\\ &= \{1056.9, 1020.5, 1202.7, 1129.8, 984.0, 1166.3, 291.6, 1640.1,\\ &\quad 1093.4, 1457.9, 1567.2, 1275.6, 328.0, 1239.2\}\end{aligned}$$

as noted previously.

Figure 7.3.3: Recording of "linear algebra rules" as displayed by Audacity



For a function f defined over the interval $[0, L]$, the (infinite) series

$$b_1 \sin\left(\frac{\pi}{L}x\right) + b_2 \sin\left(2 \cdot \frac{\pi}{L}x\right) + b_3 \sin\left(3 \cdot \frac{\pi}{L}x\right) + \cdots$$

where

$$b_m = \frac{\left\langle f, \sin\left(m \cdot \frac{\pi}{L}t\right)\right\rangle}{\left\langle \sin\left(m \cdot \frac{\pi}{L}t\right), \sin\left(m \cdot \frac{\pi}{L}t\right)\right\rangle}$$

is called the **Fourier sine series** for f. The (infinite) series

$$a_0 + a_1 \cos\left(\frac{\pi}{L}x\right) + a_2 \cos\left(2 \cdot \frac{\pi}{L}x\right) + \cdots$$

where

$$a_m = \frac{\left\langle f, \cos\left(m \cdot \frac{\pi}{L}t\right)\right\rangle}{\left\langle \cos\left(m \cdot \frac{\pi}{L}t\right), \cos\left(m \cdot \frac{\pi}{L}t\right)\right\rangle}$$

is called the **Fourier cosine series** for f. The a_m and b_m are called **Fourier coefficients**, and the functions $\sin\left(m \cdot \frac{\pi}{L}x\right)$ and $\cos\left(m \cdot \frac{\pi}{L}x\right)$ are called the m^{th} **harmonics**. Most of our effort to now has concentrated on the sine series since we have been examining functions for which $f(0) = f(L) = 0$. In general though, any piecewise continuous function can be approximated arbitrarily closely using a finite number of terms of either the Fourier sine series, as we have done, *or* the Fourier cosine series. The fit will simply be better with certain selections of harmonics.

A general Fourier series involves both sine and cosine functions and is defined over domains symmetric about zero. For any function f defined over the interval $[-L, L]$, the (infinite) series

$$a_0 + a_1 \cos\left(\frac{\pi}{L}x\right) + b_1 \sin\left(\frac{\pi}{L}x\right) + a_2 \cos\left(2 \cdot \frac{\pi}{L}x\right) + b_2 \sin\left(2 \cdot \frac{\pi}{L}x\right) + \cdots$$

where

$$a_m = \frac{\left\langle f, \cos\left(m \cdot \frac{\pi}{L}t\right)\right\rangle}{\left\langle \cos\left(m \cdot \frac{\pi}{L}t\right), \cos\left(m \cdot \frac{\pi}{L}t\right)\right\rangle} \quad \text{and} \quad b_m = \frac{\left\langle f, \sin\left(m \cdot \frac{\pi}{L}t\right)\right\rangle}{\left\langle \sin\left(m \cdot \frac{\pi}{L}t\right), \sin\left(m \cdot \frac{\pi}{L}t\right)\right\rangle}$$

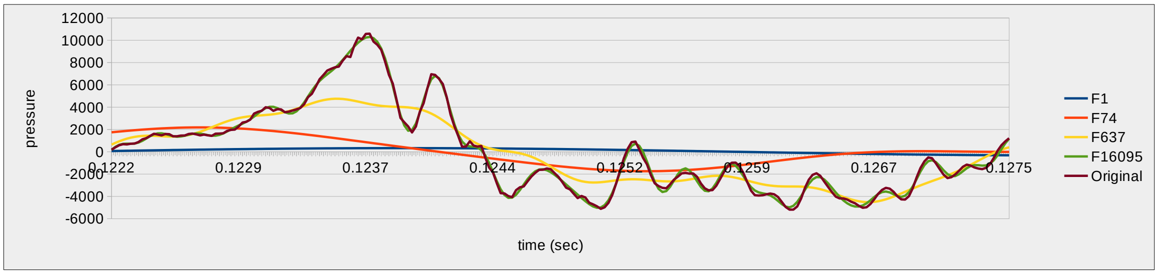
is called the **Fourier series** for f. Indeed, the term Fourier series, with no qualifiers, refers to this series, not the sine or cosine series. As before the a_m and b_m are called Fourier coefficients and the functions $\sin\left(m \cdot \frac{\pi}{L}x\right)$ and $\cos\left(m \cdot \frac{\pi}{L}x\right)$ are called the m^{th} harmonics.

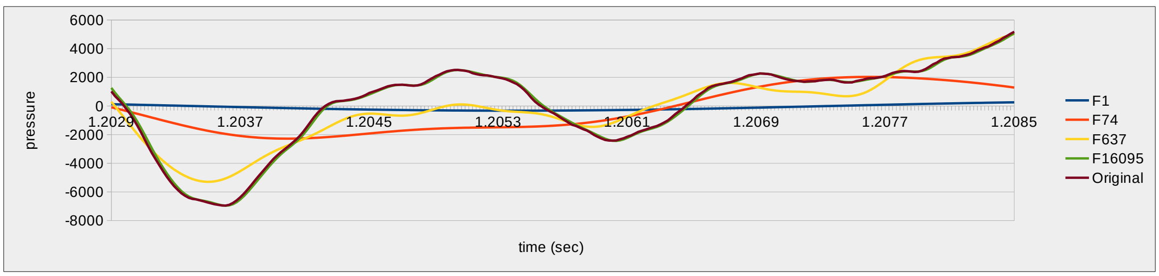
Whether any of these series converges to something useful is a deep and interesting topic of analysis. Generally, though, piecewise continuous functions can be approximated arbitrarily closely using a finite number of terms from any one of the Fourier series. Because sines and cosines are periodic, especially good approximations with small numbers of harmonics can often be found for functions f where $f(-L) = f(L)$ (for general Fourier series) or $f(0) = f(L)$ (for sine series).

It is not only functions that show some regularity or symmetry that can be approximated by Fourier series, however. The real power of Fourier analysis is to suss out the most important frequencies when none are apparent. Circling back to sound waves, figure 7.3.3[5] shows the full 1.82 seconds of a voice saying "linear algebra rules". The sound wave shows no particular symmetry or regularity since the sound is constantly changing throughout.

[5] Audacity: https://www.audacityteam.org/

The first 16,095 Fourier sine series coefficients were calculated. The following graphs of the "linear algebra rules" audio and its approximations over two separate time intervals illustrate how extending the basis improves the approximation. F1, F74, F637, and F16095 are the approximations using the dominant 1, 74, 637, and 16095 harmonics, respectively.





F16095 is barely visible beneath the original curve, suggesting that the approximation is very good (which should probably be expected having used so many terms). It may be surprising then that the distance between F16095 and the original wave is about 254.4, a number that may seem large. Distances are relative to the function being approximated, though. The norm of the original sound wave is 3493.5, so $d(\text{f16095}, \text{original})$ is only about one fourteenth (3.7%) the norm (size) of the original—not bad. On the other hand, the distances between the original and F1, F74, and F637 are 3479.5, 3032.0, and 2084.8, respectively. As their distances are similar in magnitude to the norm of the sound wave itself, it should be expected that they do a poor job approximating the original sound wave. This expectation is born out by the graphs.

In the end, though, the sound of the reproduction should be the judge of the quality of the approximation. The ancillary website contains all the data and playable sound files for the sounds mentioned in this section as well as several others, such as boiling water and birds chirping. The audio corresponding to F1 is a simple computer tone and does not resemble the original audio at all except that it captures the overall pitch. The sound of F74 is slightly better in that it oscillates, but in no way sounds like speech. The words can clearly be heard behind a noisy foreground in F637, but it is still a poor reproduction. Think Alexander Graham Bell and his first phone call. Finally, F16095 and the original audio are indistinguishable, at least to my ear. Have a listen!

Key Concepts

Fourier series for functions f defined over $[-L, L]$, the series

$$a_0 + a_1 \cos\left(\frac{\pi}{L}x\right) + b_1 \sin\left(\frac{\pi}{L}x\right) + a_2 \cos\left(2\cdot\frac{\pi}{L}x\right) + b_2 \sin\left(2\cdot\frac{\pi}{L}x\right) + \cdots$$

where

$$a_m = \frac{\left\langle f, \cos\left(m\cdot\frac{\pi}{L}t\right)\right\rangle}{\left\langle \cos\left(m\cdot\frac{\pi}{L}t\right), \cos\left(m\cdot\frac{\pi}{L}t\right)\right\rangle} \quad \text{and} \quad b_m = \frac{\left\langle f, \sin\left(m\cdot\frac{\pi}{L}t\right)\right\rangle}{\left\langle \sin\left(m\cdot\frac{\pi}{L}t\right), \sin\left(m\cdot\frac{\pi}{L}t\right)\right\rangle}.$$

Fourier sine series for functions f defined over $[0, L]$, the series

$$b_1 \sin\left(\frac{\pi}{L}x\right) + b_2 \sin\left(2\cdot\frac{\pi}{L}x\right) + \cdots$$

where

$$b_m = \frac{\left\langle f, \sin\left(m \cdot \frac{\pi}{L} t\right)\right\rangle}{\left\langle \sin\left(m \cdot \frac{\pi}{L} t\right), \sin\left(m \cdot \frac{\pi}{L} t\right)\right\rangle}.$$

Fourier cosine series for functions f defined over $[0, L]$, the series

$$a_0 + a_1 \cos\left(\frac{\pi}{L} x\right) + a_2 \cos\left(2 \cdot \frac{\pi}{L} x\right) + \cdots$$

where

$$a_m = \frac{\left\langle f, \cos\left(m \cdot \frac{\pi}{L} t\right)\right\rangle}{\left\langle \cos\left(m \cdot \frac{\pi}{L} t\right), \cos\left(m \cdot \frac{\pi}{L} t\right)\right\rangle}.$$

Fourier coefficients the a_m and b_m of a Fourier series, Fourier sine series, or Fourier cosine series.

harmonics the functions $\cos\left(m \cdot \frac{\pi}{L} t\right)$, and $\sin\left(m \cdot \frac{\pi}{L} t\right)$ appearing in Fourier series are called m^{th} harmonics, or simply harmonics when not referring to any particular frequency.

approximations piecewise continuous functions can be approximated arbitrarily closely using a finite number of terms from any one of the Fourier series. Especially good approximations with small numbers of harmonics can often be found for smooth functions f where $f(-L) = f(L)$ (for general Fourier series) or $f(0) = f(L)$ (for sine or cosine series).

Fourier analysis the process of determining a large selection of Fourier coefficients with the purpose of identifying those with some particular characteristic.

Exercises

1. Argue that $C([0, L])$ is a vector space. [𝔸]-363
2. Justify (7.3.3).
3. Why do we not include the case $m = 0$ in the family of harmonics for Fourier sine series? [𝔸]-363
4. Show that, in the inner product space $C([0, L])$,
   (a) $\langle 1, 1\rangle = L$
   (b) $\left\langle \cos\left(m \cdot \frac{\pi}{L} t\right), \cos\left(m \cdot \frac{\pi}{L} t\right)\right\rangle = \frac{L}{2}$ for $m = 1, 2, \ldots$ [𝕊]-339
   (c) $\left\langle \sin\left(m \cdot \frac{\pi}{L} t\right), \sin\left(m \cdot \frac{\pi}{L} t\right)\right\rangle = \frac{L}{2}$ for $m = 1, 2, \ldots$
5. Find the Fourier sine series for f over the given interval. Use symmetry whenever possible to help with the calculations.
   (a) $f(x) = 1$; $[0, 1]$
   (b) $f(x) = x$; $[0, 1]$ [𝕊]-339
   (c) $f(x) = \begin{cases} x & x \le 1/2 \\ 1 - x & x > 1/2 \end{cases}$; $[0, 1]$
   (d) $f(x) = e^x$; $[0, \ln 2]$ [𝔸]-363
6. Find the Fourier cosine series for f over the given interval. Use symmetry whenever possible to help with the calculations.
   (a) $f(x) = \frac{1}{2} - x$; $[0, 1]$
   (b) $f(x) = x$; $[0, 1]$ [𝕊]-339
   (c) $f(x) = \begin{cases} x & x \le 1/2 \\ 1 - x & x > 1/2 \end{cases}$; $[0, 1]$
   (d) $f(x) = e^x$; $[0, \ln 2]$ [𝔸]-363
7. Find the Fourier series for f over the interval $[-1, 1]$.
   (a) $f(x) = 1$ [𝕊]-340
   (b) $f(x) = x$
   (c) $f(x) = |x|$
   (d) $f(x) = e^x$
   (e) $f(x) = \frac{|x|}{x}$
8. Create graphs of f and the Fourier series of question 5 with the first (i) 1 nonzero term; (ii) 2 nonzero terms; and (iii) 5 nonzero terms. [𝔸]-363
9. Create graphs of f and the Fourier series of question 6 with the first (i) 1 nonzero term; (ii) 2 nonzero terms; and (iii) 5 nonzero terms. [𝔸]-364
10. Create graphs of f and the Fourier series of question 7 with the first (i) 1 nonzero term; (ii) 3 nonzero terms; and (iii) 5 nonzero terms. [𝔸]-365
11. Reproduce a sound wave, part 1.
    (a) Download one of the data spreadsheets of the sounds on the ancillary website.
    (b) Sort the Fourier sine series coefficients by decreasing magnitude.
    (c) Using the twenty harmonics with greatest magnitude coefficients, reproduce the sound wave as a sum of these twenty sine functions.

(d) Graph the original sound wave and its reproduction on the same set of axes.

12. Reproduce a sound wave, part 2.

(a) Grab about 600 samples (6/441 sec) from one of the data spreadsheets of the sounds on the ancillary website.

(b) Compute the Fourier cosine series coefficients for the first 100 harmonics. You will need to write a computer program that implements a numerical integration technique. The trapezoidal rule will suffice, for example.

(c) Using the twenty harmonics with greatest magnitude coefficients, reproduce the sound wave as a sum of cosine functions.

(d) Graph the original sound wave and its reproduction on the same set of axes.

Answers

inner product on $C([0, L])$ The properties of an inner product are justified one by one below.

1. For any function f in $C([0, L])$, $\langle f, f\rangle = \int_0^L f^2(x)\,dx \geq 0$ since $f^2(x) \geq 0$ for all x in $[0, L]$. In other words, $f^2(x)$ is nonnegative, so its definite integral is nonnegative.
2. Of course, if $f(x) = 0$ (that is, $f = \mathbf{0}$), then $\langle f, f\rangle = \int_0^L f^2(x)\,dx = \int_0^L 0\,dx = 0$. Now suppose $f \neq \mathbf{0}$. That is, there is some x_0 in $[0, L]$ for which $f(x_0) \neq 0$. Let $z = f^2(x_0) > 0$. Since f^2 is continuous, there is a δ such that whenever $|x - x_0| < \delta$, x in $[0, L]$, $|f^2(x) - f^2(x_0)| = |f^2(x) - z| < \frac{z}{2}$. This establishes an interval I of width at least δ within $[0, L]$ where $f^2(x) > \frac{z}{2}$ so $\langle f, f\rangle = \int_0^L f^2(x)\,dx \geq \int_I f^2(x)\,dx \geq \delta\frac{z}{2} > 0$. Hence if $f \neq \mathbf{0}$ then $\langle f, f\rangle \neq 0$, or contrapositively if $\langle f, f\rangle = 0$ then $f = \mathbf{0}$.
3. For any f, g in $C([0, L])$, $\langle f, g\rangle = \int_0^L f(x)g(x)\,dx = \int_0^L g(x)f(x)\,dx = \langle g, f\rangle$ since multiplication is commutative.
4. For any f, g, h in $C([0, L])$,
$$\begin{aligned}\langle f + g, h\rangle &= \int_0^L (f(x) + g(x))h(x)\,dx = \int_0^L (f(x)h(x) + g(x)h(x))\,dx\\ &= \int_0^L f(x)h(x)\,dx + \int_0^L g(x)h(x)\,dx = \langle f, h\rangle + \langle g, h\rangle\end{aligned}$$
by the distributive property for real numbers and a standard result of calculus.
5. For any f, g in $C([0, L])$ and any scalar c, $\langle cf, g\rangle = \int_0^L cf(x)g(x)\,dx = c\int_0^L f(x)g(x)\,dx = c\langle f, g\rangle$ by a standard result of calculus.

sine functions are orthogonal By (7.3.4),

$$\sin\left(\frac{m\pi}{L}t\right)\sin\left(\frac{n\pi}{L}t\right) = \frac{1}{2}\left[\cos\left((m-n)\frac{\pi}{L}t\right) - \cos\left((m+n)\frac{\pi}{L}t\right)\right]$$

so

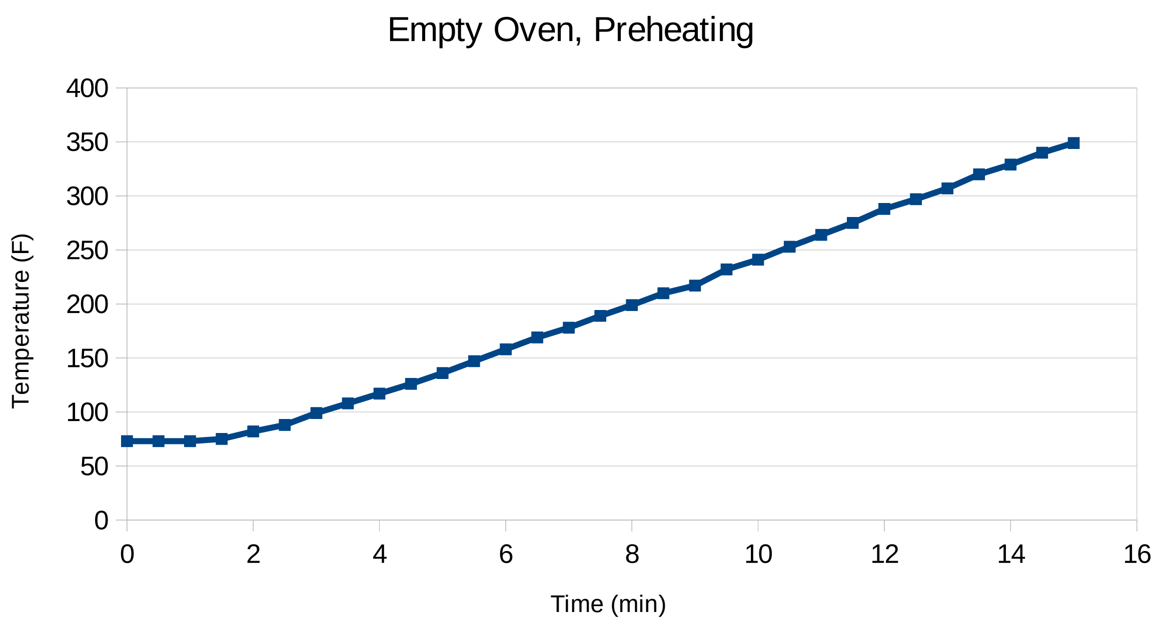
$$\begin{aligned}\left\langle \sin\left(\frac{m\pi}{L}t\right), \sin\left(\frac{n\pi}{L}t\right)\right\rangle &= \int_0^L \sin\left(\frac{m\pi}{L}t\right)\sin\left(\frac{n\pi}{L}t\right)dt\\ &= \frac{1}{2}\int_0^L \left[\cos\left((m-n)\frac{\pi}{L}t\right) - \cos\left((m+n)\frac{\pi}{L}t\right)\right]dt\\ &= \frac{1}{2}\left[\frac{L}{\pi(m-n)}\sin\left((m-n)\frac{\pi}{L}t\right) - \frac{L}{\pi(m+n)}\sin\left((m+n)\frac{\pi}{L}t\right)\right]_0^L\\ &= 0.\end{aligned}$$

7.4 Discrete Dynamical Systems [3.2, 3.3]

You have just finished chopping, slicing, mixing, blending, marinating, layering, and otherwise preparing your favorite dish. You are ready to place it in the oven when you realize it has not been preheated. Preheat now, or put your assembled dish in, start the oven and guess how long it will take to properly bake? Neither! Model the situation with a discrete dynamical system and know just how long to put it in a cold oven.

After doing the experiment once, you will know just what to do next time you forget to preheat. For the experiment, place a thermometer probe in the empty oven approximately where you will later put the brownies. Note the temperature of the probe (air inside the oven) and begin preheating the oven. Record the thermometer reading every 30 seconds until the oven reaches the target temperature (likely 350°F/175°C for brownies but you may want to coninue to higher temperature if you are baking something else).

Make a graph of temperature versus time. If your oven is like mine, you will get a curve that looks something like this.[6]



Curiously there is essentially no heating during the first 90 seconds, after which the temperature increases in a very steadily linear fashion at about 21°F per minute. We will use this observation to model the temperature of the oven during preheating.

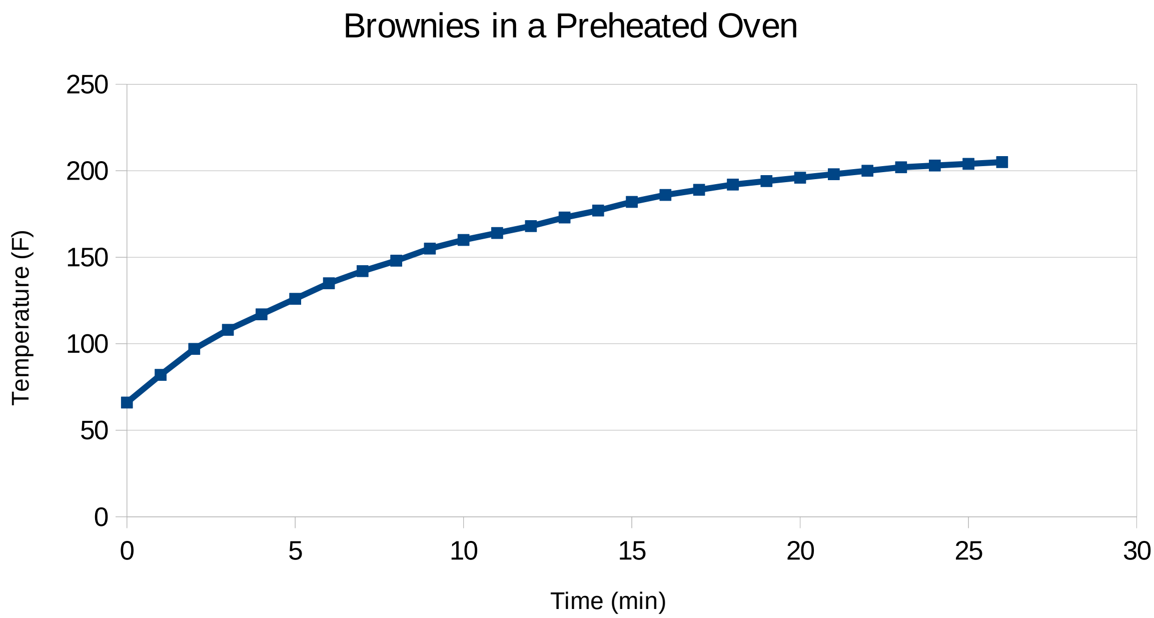
Remove the probe from the oven, let it cool, and let the oven return to its preheated temperature. When the probe has cooled to room temperature and the brownie mix has been poured into the brownie pan, insert the probe into the brownie mix. Record its internal temperature and put the brownies in the preheated oven. Record the thermometer reading every minute until the brownies are done. You will notice the heating is not linear.

Newton's law of cooling, which applies equally to heating, suggests that the change in temperature of a body is approximately proportional to the difference between the temperature of the body and the temperature of its surroundings, ambient temperature. As an equation,

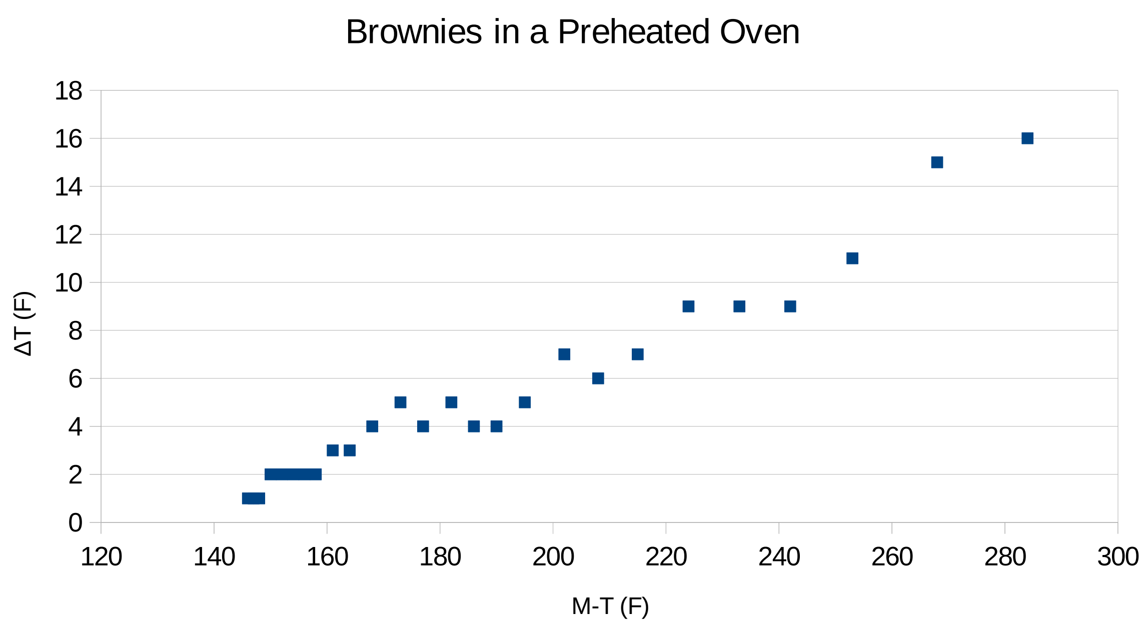
$$\Delta T \approx k(M - T).$$

If brownies obey this law, plotting the temperature over time will reveal a concave down graph. As the brownies' temperature increases, the difference between ambient (oven) temperature and brownie temperature, $(M - T)$, decreases. In turn, the change in temperature over a fixed amount of time, ΔT, will also decrease. This is, at least as a general characteristic, exactly what the data provide!

[6]Data, graphs, and calculations for this entire discussion are available in a spreadsheet at the ancillary website.



However, if the brownies truly follow Newton's law of cooling, a plot of $M - T$ versus ΔT will reveal a straight line passing through the origin, just as any two directly related variables will. Alas this is not what the data suggest.



The scatterplot looks quite linear, but clearly would not pass through the origin if extended to $M - T = 0$. Unfortunately, this is a critical feature of the law. When there is no difference between the temperature of a body and its surroundings, the body will neither heat nor cool. Laying a glass of water on the counter for hours, days, or weeks, it will remain at room temperature for the duration.

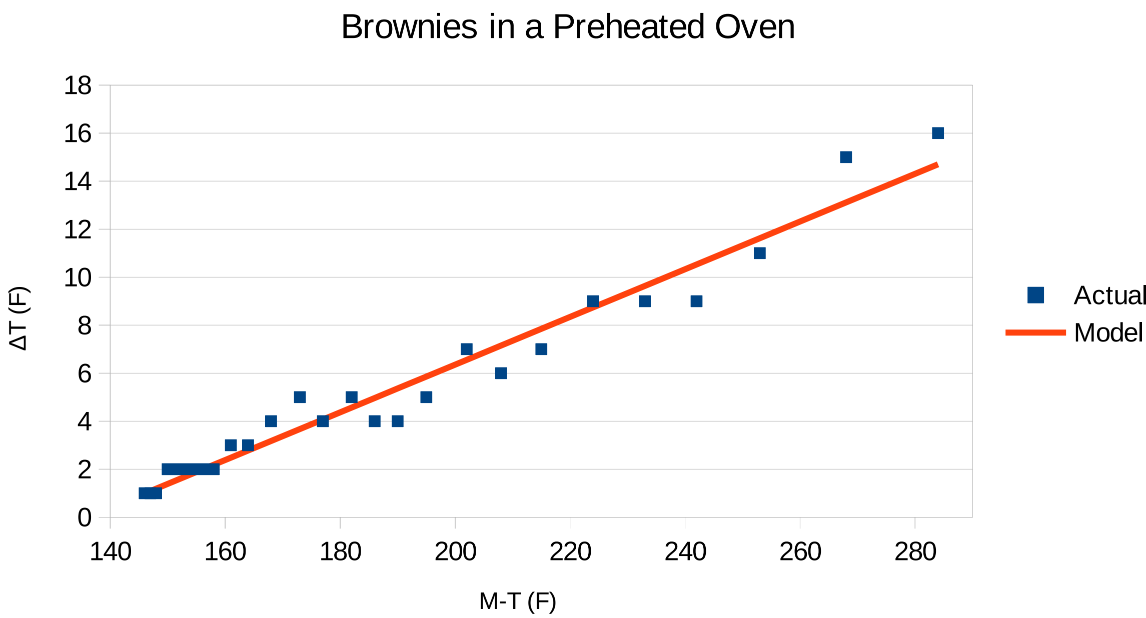
Nonetheless, this is what the data are telling us, law or no law. We apply linear regression (section 7.1) to the data, deriving a model of the from $\Delta T = \alpha_0 + \alpha_1(M - T)$ that applies when $146°\text{F} \leq M - T \leq 284°\text{F}$. The normal equations are

$$\begin{bmatrix} 26 & 4936 \\ 4936 & 977404 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 139 \\ 30395 \end{bmatrix}$$

and have solution

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 26 & 4936 \\ 4936 & 977404 \end{bmatrix}^{-1} \begin{bmatrix} 139 \\ 30395 \end{bmatrix} = \begin{bmatrix} -13.516 \\ .099356 \end{bmatrix}.$$

Hence the temperature of the brownies can reasonably be modeled by $\Delta T = -13.516 + .099356(M - T)$. The graph here illustrates the reasonableness of this model.



This model looks very good for temperature differennces, $M - T$, above 146°F, but putting brownies in a cold oven requires a model for much smaller temperature differences. After all, the brownie mix and the unheated oven are essentially the same temperature to begin, $M - T \approx 0$. Luckily Newton's law of cooling applies for "small" temperature differences. We can conclude that for our brownies the 350°F oven provides a temperature difference outside the range of "small". Without any further data, we will assume that Newton's law of cooling applies at temperature differences less than 146°F (the smallest observed temperature difference). According to our model, at $M - T = 146$, we have $\Delta T = -13.516 + .099356(146) = 0.98998 \approx 1$. Since Newton's law implies that ΔT and $M - T$ are directly proportional, we arrive at the simple relation $\Delta T = \frac{1}{146}(M - T)$ for $0 \le M - T \le 146$.

Finally we are ready to run a simulation and find out just how long the brownies should bake starting in a cold oven. From the observation of oven temperature during preheating, we have

$$\Delta M = 21$$

90 seconds or more into heating (and $\Delta M = 0$ prior since the oven does not heat during the first 90 seconds). Substituting $\Delta T = T(t+1) - T(t)$ and $\Delta M = M(t+1) - M(t)$, the changes in temperature over the course of 1 minute, we have starting at 1.5 minutes,

$$\begin{bmatrix} M(t+1) \\ T(t+1) \end{bmatrix} = \begin{bmatrix} M(t) \\ T(t) \end{bmatrix} + \begin{bmatrix} 21 \\ \frac{1}{146}(M(t) - T(t)) \end{bmatrix}. \tag{7.4.1}$$

Given that the brownies began at 66°F and the oven began at 72°F, we have

$$\begin{aligned}
\begin{bmatrix} M(1.5) \\ T(1.5) \end{bmatrix} &= \begin{bmatrix} 72 \\ 66 \end{bmatrix}, \\
\begin{bmatrix} M(2.5) \\ T(2.5) \end{bmatrix} &= \begin{bmatrix} M(1.5) \\ T(1.5) \end{bmatrix} + \begin{bmatrix} 21 \\ \frac{1}{146}(M(1.5) - T(1.5)) \end{bmatrix} = \begin{bmatrix} 72 \\ 66 \end{bmatrix} + \begin{bmatrix} 21 \\ \frac{1}{146}(72 - 66) \end{bmatrix} = \begin{bmatrix} 93 \\ 66.041 \end{bmatrix}, \\
\begin{bmatrix} M(3.5) \\ T(3.5) \end{bmatrix} &= \begin{bmatrix} M(2.5) \\ T(2.5) \end{bmatrix} + \begin{bmatrix} 21 \\ \frac{1}{146}(M(2.5) - T(2.5)) \end{bmatrix} = \begin{bmatrix} 114 \\ 66.226 \end{bmatrix}, \ldots
\end{aligned}$$

and more succinctly,

$$\begin{bmatrix} M(1.5) \\ T(1.5) \end{bmatrix}, \begin{bmatrix} M(2.5) \\ T(2.5) \end{bmatrix}, \begin{bmatrix} M(3.5) \\ T(3.5) \end{bmatrix}, \ldots =$$

$$\begin{bmatrix} 72 \\ 66 \end{bmatrix}, \begin{bmatrix} 93 \\ 66.041 \end{bmatrix}, \begin{bmatrix} 114 \\ 66.226 \end{bmatrix}, \begin{bmatrix} 135 \\ 66.553 \end{bmatrix}, \begin{bmatrix} 156 \\ 67.022 \end{bmatrix}, \begin{bmatrix} 177 \\ 67.631 \end{bmatrix}, \begin{bmatrix} 198 \\ 68.380 \end{bmatrix}, \begin{bmatrix} 219 \\ 69.268 \end{bmatrix}, \ldots$$

bringing us to 8.5 minutes. At this point, $M - T = 219 - 69.268 = 149.73$, which exceeds 146. To continue, we need to start using $\Delta T = -13.516 + .099356(M - T)$, or $T(t+1) = T(t) - 13.516 + .099356(M(t) - T(t))$ for the change in brownie temperature. In other words, we now have

$$\begin{bmatrix} M(t+1) \\ T(t+1) \end{bmatrix} = \begin{bmatrix} M(t) \\ T(t) \end{bmatrix} + \begin{bmatrix} 21 \\ -13.516 + .099356(M(t) - T(t)) \end{bmatrix}. \tag{7.4.2}$$

So

$$\begin{bmatrix} M(9.5) \\ T(9.5) \end{bmatrix} = \begin{bmatrix} 219 \\ 69.268 \end{bmatrix} + \begin{bmatrix} 21 \\ -13.516 + .099356(149.73) \end{bmatrix} = \begin{bmatrix} 240 \\ 70.629 \end{bmatrix}$$

and so on,

$$\begin{bmatrix} M(10.5) \\ T(10.5) \end{bmatrix}, \begin{bmatrix} M(11.5) \\ T(11.5) \end{bmatrix}, \begin{bmatrix} M(12.5) \\ T(12.5) \end{bmatrix}, \ldots =$$
$$\begin{bmatrix} 261 \\ 73.941 \end{bmatrix}, \begin{bmatrix} 282 \\ 79.010 \end{bmatrix}, \begin{bmatrix} 303 \\ 85.662 \end{bmatrix}, \begin{bmatrix} 324 \\ 93.740 \end{bmatrix}, \begin{bmatrix} 345 \\ 103.10 \end{bmatrix}, \ldots$$

at which point we reach another milestone. The oven temperature does not jump another 21°F at this point. It will only increase another 5°F, so is essentially up to working temperature. The first 14.5 minutes of baking brings the oven to ~ 350°F and the brownies to ~ 103°F, a state that the brownies baking in a preheated oven reached in about 2.5 minutes. To summarize, brownies in a cold oven took 14.5 minutes to get to the same point (oven temperature 350°F, brownie temperature 103°F) the brownies reached in a preheated oven in only 2.5 minutes. From here out it is safe to assume the baking will proceed similarly. Therefore it takes 12 more minutes to bake brownies starting in a cold oven than it does starting in a preheated oven. We simply add 12 minutes to the baking time and proceed. Presumably this applies to any baking done at 350°F. The first 14.5 minutes of baking starting with a cold oven are equivalent to only 2.5 minutes of baking starting with a preheated oven.

As fascinating as the brownie heating experiment may be, this is neither an engineering nor math modeling class. Not to worry, the reader will not be asked to create their own models. Instead, focus on the results of the modeling process, equations (7.4.1) and (7.4.2). These are discrete dynamical systems. The talk about brownies and ovens has hopefuly grabbed your attention and motivated study, nothing more. In case not even that, I should mention discrete dynamical systems are used to model phenomena in biology, medicine, physics, economics, engineering, and a host of other areas. Chances are, if you are studying linear algebra, discrete dynamical systems appear in your field of study.

A **first order discrete dynamical system** is an equation

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k), \tag{7.4.3}$$

which paired with an **initial condition**

$$\mathbf{x}_0 = \mathbf{v} \tag{7.4.4}$$

defines a sequence $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$. (7.4.3) is an example of a **recurrence** or **recurrence relation** and determines all of the terms of the sequence except the first, which must be supplied separately. For purpose of our study of linear algebra, $\mathbf{x}_k$ are in $\mathbb{R}^n$ and $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is an arbitrary function.

Equation (7.4.1) can be rewritten in terms of this definition by setting $\mathbf{x}_k = \begin{bmatrix} M(k) \\ T(k) \end{bmatrix}$, from which it follows

$$\begin{aligned}
\mathbf{f}(\mathbf{x}_k) &= \begin{bmatrix} M(k) \\ T(k) \end{bmatrix} + \begin{bmatrix} 21 \\ \frac{1}{146}(M(k) - T(k)) \end{bmatrix} \\
&= \begin{bmatrix} M(k) \\ T(k) \end{bmatrix} + \frac{1}{146}\begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} M(k) \\ T(k) \end{bmatrix} + \begin{bmatrix} 21 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} M(k) \\ T(k) \end{bmatrix} + \frac{1}{146}\begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} M(k) \\ T(k) \end{bmatrix} + \begin{bmatrix} 21 \\ 0 \end{bmatrix} \\
&= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{146}\begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}\right)\begin{bmatrix} M(k) \\ T(k) \end{bmatrix} + \begin{bmatrix} 21 \\ 0 \end{bmatrix} \\
&= \frac{1}{146}\begin{bmatrix} 146 & 0 \\ 1 & 145 \end{bmatrix}\begin{bmatrix} M(k) \\ T(k) \end{bmatrix} + \begin{bmatrix} 21 \\ 0 \end{bmatrix} \\
&= \frac{1}{146}\begin{bmatrix} 146 & 0 \\ 1 & 145 \end{bmatrix}\mathbf{x}_k + \begin{bmatrix} 21 \\ 0 \end{bmatrix}.
\end{aligned}$$

This is an example of a **nonlinear discrete dynamical system** as the function $\mathbf{f}$ is not a linear transformation. In this case, the function $\mathbf{f}$ is **affine**, and we would therefore say the system is affine. It takes the form

$$\mathbf{x}_{k+1} = M\mathbf{x}_k + \mathbf{b}. \tag{7.4.5}$$

As noted in the definition, once an initial condition is provided, a discrete dynamical system determines a sequence. Our first order of business is to understand how so. For each initial condition, the sequence defined by a discrete dynamical system can be calculated term-by-term. The recurrence relation defines each term after the first. For example, the first few terms of the sequence defined by the system

$$\mathbf{x}_{k+1} = \begin{bmatrix} -.15 & -1.3 & -1.95 \\ -.55 & 1.8 & 3.15 \\ .15 & -1.3 & -2.25 \end{bmatrix}\mathbf{x}_k + \begin{bmatrix} 2 \\ -5 \\ 4 \end{bmatrix} \tag{7.4.6}$$

with initial condition

$$\mathbf{x}_0 = \begin{bmatrix} 8 \\ -5 \\ 6 \end{bmatrix}$$

can be calculated as follows. According to (7.4.6),

$$\mathbf{x}_1 = \begin{bmatrix} -.15 & -1.3 & -1.95 \\ -.55 & 1.8 & 3.15 \\ .15 & -1.3 & -2.25 \end{bmatrix}\mathbf{x}_0 + \begin{bmatrix} 2 \\ -5 \\ 4 \end{bmatrix} = \begin{bmatrix} -.15 & -1.3 & -1.95 \\ -.55 & 1.8 & 3.15 \\ .15 & -1.3 & -2.25 \end{bmatrix}\begin{bmatrix} 8 \\ -5 \\ 6 \end{bmatrix} + \begin{bmatrix} 2 \\ -5 \\ 4 \end{bmatrix} = \begin{bmatrix} -4.4 \\ .5 \\ -1.8 \end{bmatrix}.$$

Also according to (7.4.6),

$$\mathbf{x}_2 = \begin{bmatrix} -.15 & -1.3 & -1.95 \\ -.55 & 1.8 & 3.15 \\ .15 & -1.3 & -2.25 \end{bmatrix}\mathbf{x}_1 + \begin{bmatrix} 2 \\ -5 \\ 4 \end{bmatrix} = \begin{bmatrix} -.15 & -1.3 & -1.95 \\ -.55 & 1.8 & 3.15 \\ .15 & -1.3 & -2.25 \end{bmatrix}\begin{bmatrix} -4.4 \\ .5 \\ -1.8 \end{bmatrix} + \begin{bmatrix} 2 \\ -5 \\ 4 \end{bmatrix} = \begin{bmatrix} 5.52 \\ -7.35 \\ 6.74 \end{bmatrix}.$$

Similarly,

$$\mathbf{x}_3 = \begin{bmatrix} -.15 & -1.3 & -1.95 \\ -.55 & 1.8 & 3.15 \\ .15 & -1.3 & -2.25 \end{bmatrix}\begin{bmatrix} 5.52 \\ -7.35 \\ 6.74 \end{bmatrix} + \begin{bmatrix} 2 \\ -5 \\ 4 \end{bmatrix} = \begin{bmatrix} -2.416 \\ -.035 \\ -.782 \end{bmatrix}$$

and

$$\mathbf{x}_4 = \begin{bmatrix} -.15 & -1.3 & -1.95 \\ -.55 & 1.8 & 3.15 \\ .15 & -1.3 & -2.25 \end{bmatrix}\begin{bmatrix} -2.416 \\ -.035 \\ -.782 \end{bmatrix} + \begin{bmatrix} 2 \\ -5 \\ 4 \end{bmatrix} = \begin{bmatrix} 3.933 \\ -6.198 \\ 5.443 \end{bmatrix}$$

accurate to 3 decimal places. The first five terms of the sequence are $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$, which have just been calculated (using this SageCell) as

$$\begin{bmatrix} 8 \\ -5 \\ 6 \end{bmatrix}, \begin{bmatrix} -4.4 \\ .5 \\ -1.8 \end{bmatrix}, \begin{bmatrix} 5.52 \\ -7.35 \\ 6.74 \end{bmatrix}, \begin{bmatrix} -2.416 \\ -.035 \\ -.782 \end{bmatrix}, \begin{bmatrix} 3.933 \\ -6.198 \\ 5.443 \end{bmatrix}.$$

Further terms can be calculated similarly. The process of calculating the terms is called **iteration**. The terms themselves are called **iterates** or **iterations**, and the sequence is called the **orbit** of $\mathbf{x}_0$.

Given the dynamical system

$$\mathbf{x}_{k+1} = \frac{1}{146}\begin{bmatrix} 146 & 0 \\ 1 & 145 \end{bmatrix}\mathbf{x}_k + \begin{bmatrix} 21 \\ 0 \end{bmatrix} \tag{7.4.7}$$

from the brownie baking model, can you find the first 5 iterates in the orbit of

$$\mathbf{x}_0 = \begin{bmatrix} 72 \\ 66 \end{bmatrix}?$$

Answer on page 267.

As a final exercise in iteration, the first 5 iterates in the orbit of $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for the dynamical system

$$\mathbf{x}_{k+1} = \begin{bmatrix} \frac{\sqrt{6}+\sqrt{2}}{4} & \frac{-\sqrt{6}+\sqrt{2}}{4} \\ \frac{\sqrt{6}-\sqrt{2}}{4} & \frac{\sqrt{6}+\sqrt{2}}{4} \end{bmatrix}\mathbf{x}_k + \begin{bmatrix} -2 \\ 2 \end{bmatrix} \tag{7.4.8}$$

are (approximately)

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1.293 \\ 3.224 \end{bmatrix}, \begin{bmatrix} -4.083 \\ 4.780 \end{bmatrix}, \begin{bmatrix} -7.182 \\ 5.560 \end{bmatrix}, \begin{bmatrix} -10.376 \\ 5.512 \end{bmatrix}.$$

Can you verify these terms (using SageMath)? Answer on page 267.

The first few iterates of an orbit are often not the ultimate goal, however. For many applications, the point of interest is the long run. How can the 1000^{th} through 2000^{th} or $1,000,000^{th}$ through $1,000,012^{th}$ iterations of an orbit be described in general terms? Such a desctiption is called the system's **long term behavior**.

In the case of (7.4.6),

$$\mathbf{x}_5 = \begin{bmatrix} -1.146 \\ -1.174 \\ 0.401 \end{bmatrix}, \mathbf{x}_{30} = \begin{bmatrix} 1.108 \\ -3.416 \\ 2.647 \end{bmatrix}, \mathbf{x}_{55} = \begin{bmatrix} 1.111 \\ -3.419 \\ 2.650 \end{bmatrix}, \text{ and } \mathbf{x}_{80} = \begin{bmatrix} 1.111 \\ -3.419 \\ 2.650 \end{bmatrix}$$

accurate to 3 decimal places. It takes a short while, but the terms reveal a pattern. All terms $\mathbf{x}_k$, $k \geq 55$ are, accurate to 3 decimal places, equal to $\begin{bmatrix} 1.111 \\ -3.419 \\ 2.650 \end{bmatrix}$. The iterates after the 54^{th} do not change much. When the iterates of a dynamical system settle down this way for all initial values in some neighborhood, we say that the vector it is settling on is an **attractor**. It "pulls" orbits toward it. But what vector is the attractor, and can we predict it without computing large numbers of iterates? By definition, $\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k)$, so when a dynamical system has an attractor, it means $\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k) \approx \mathbf{x}_k$ and the approximation improves as k increases. Therein lies the answer to the mystery.

The iterates are getting closer and closer to satisfying the equation

$$\mathbf{f}(\mathbf{x}) = \mathbf{x}. \tag{7.4.9}$$

Any solution of this equation is called a **fixed point** of $\mathbf{f}$, and if $\mathbf{x}_k$ were such a value, we would have $\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k) = \mathbf{x}_k$. The sequence would be fixed forever more at the value $\mathbf{x}$.

For example, a fixed point of (7.4.6) satisfies

$$\begin{bmatrix} -.15 & -1.3 & -1.95 \\ -.55 & 1.8 & 3.15 \\ .15 & -1.3 & -2.25 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 2 \\ -5 \\ 4 \end{bmatrix} = \mathbf{x},$$

an equation we can solve:

$$\begin{bmatrix} -.15 & -1.3 & -1.95 \\ -.55 & 1.8 & 3.15 \\ .15 & -1.3 & -2.25 \end{bmatrix}\mathbf{x} - \mathbf{x} = -\begin{bmatrix} 2 \\ -5 \\ 4 \end{bmatrix}$$

$$\left(\begin{bmatrix} -.15 & -1.3 & -1.95 \\ -.55 & 1.8 & 3.15 \\ .15 & -1.3 & -2.25 \end{bmatrix} - I\right)\mathbf{x} = -\begin{bmatrix} 2 \\ -5 \\ 4 \end{bmatrix}.$$

By row reduction (using this SageCell), it turns out

$$\mathbf{x} = \frac{1}{117}\begin{bmatrix} 130 \\ -400 \\ 310 \end{bmatrix} \approx \begin{bmatrix} 1.111111111 \\ -3.418803419 \\ 2.649572650 \end{bmatrix}.$$

It seems clear enough the orbit is **approaching the fixed point**, so we have that

$$\begin{bmatrix} 1.111111111 \\ -3.418803419 \\ 2.649572650 \end{bmatrix}$$

is an attractor of (7.4.6).

Not all orbits of discrete dynamical systems approach a fixed point, however. The dynamical system

$$\mathbf{x}_{k+1} = \frac{1}{146}\begin{bmatrix} 146 & 0 \\ 1 & 145 \end{bmatrix}\mathbf{x}_k + \begin{bmatrix} 21 \\ 0 \end{bmatrix}$$

with initial condition

$$\mathbf{x}_0 = \begin{bmatrix} 72 \\ 66 \end{bmatrix}$$

from the brownie baking model does not. This fact is clear by observing the behavior of the first entry of $\mathbf{x}_k$. It simply increases by 21 with each iteration. To be precise, $(\mathbf{x}_k)_{1,1} = 72 + 21k$, which tends to infinity as k grows. Therefore, $||\mathbf{x}_k||$ diverges to ∞ and we say the orbit **tends toward infinity**.

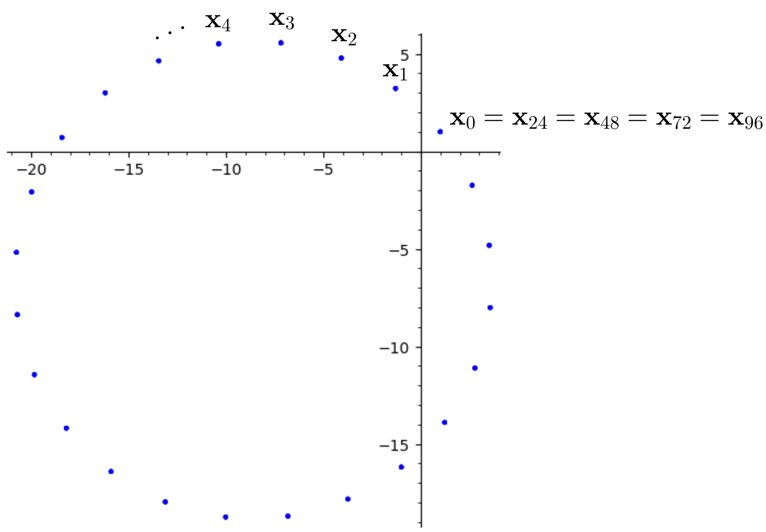
Finally, the orbit of $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for the dynamical system

$$\mathbf{x}_{k+1} = \begin{bmatrix} \frac{\sqrt{6}+\sqrt{2}}{4} & \frac{-\sqrt{6}+\sqrt{2}}{4} \\ \frac{\sqrt{6}-\sqrt{2}}{4} & \frac{\sqrt{6}+\sqrt{2}}{4} \end{bmatrix}\mathbf{x}_k + \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

does not present a clear pattern even after 100 iterations. As computed in this SageCell, $\mathbf{x}_{98}$ through $\mathbf{x}_{100}$ are (again accurate to three decimal places),

$$\begin{bmatrix} -4.083 \\ 4.780 \end{bmatrix}, \begin{bmatrix} -7.182 \\ 5.560 \end{bmatrix}, \begin{bmatrix} -10.376 \\ 5.512 \end{bmatrix}.$$

Though it is likely not at all clear from this short list nor the SageMath output, the orbit exhibits a very simple pattern. To see it, a graph of the first 100 iterations:



It appears there are only 24 iterations (points on the graph), but that is because they repeat. As indicated, $\mathbf{x}_0 = \mathbf{x}_{24} = \mathbf{x}_{48} = \cdots$. Similarly, $\mathbf{x}_1 = \mathbf{x}_{25} = \mathbf{x}_{49} = \cdots$ and so on. As a result, $\mathbf{x}_{98}$ coincides with $\mathbf{x}_2$, $\mathbf{x}_{99}$ coincides with $\mathbf{x}_3$, and $\mathbf{x}_{100}$ coincides with $\mathbf{x}_4$. When a sequence of iterates repeats this way, we say the orbit is **periodic**. A sequence of iterates that approaches such a repeating sequence is called **asymptotically periodic**.

As with the power method (section 6.2) and Markov chains (section 7.2), both of which can be framed as discrete dynamical systems, eigenvalues tell the story of long term behavior. For the example systems of this section, each of the form (7.4.5), eigenvalues of M and its spectral radius are listed in the chart.

System	**Eigenvalues of M**	**Spectral Radius**	**Long term behavior**
(7.4.6)	$-.8, -.3, .5$	.8	approaches fixed point
(7.4.7)	$1, \frac{145}{146}$	1	tends toward infinity
(7.4.8)	$\frac{\sqrt{6}+\sqrt{2}}{4} \pm i\frac{\sqrt{6}-\sqrt{2}}{4}$	1	periodic

The **spectral radius** of a square matrix is the maximum of the magnitudes (absolute values) of its eigenvalues. The magnitude of a complex number $a + ib$ is $\sqrt{a^2 + b^2}$, so

$$\begin{aligned}
\left|\frac{\sqrt{6}+\sqrt{2}}{4} \pm i\frac{\sqrt{6}-\sqrt{2}}{4}\right| &= \sqrt{\left(\frac{\sqrt{6}+\sqrt{2}}{4}\right)^2 + \left(\frac{\sqrt{6}-\sqrt{2}}{4}\right)^2} \\
&= \sqrt{\frac{6+2\sqrt{12}+2}{16} + \frac{6-2\sqrt{12}+2}{16}} \\
&= \sqrt{\frac{6+2+6+2}{16}} \\
&= 1.
\end{aligned}$$

Much like a geometric series, which converges when the ratio between consecutive terms is less than one and diverges when the ratio is greater than one, an affine system will approach the fixed point when the spectral radius is less than one and will tend toward infinity when the spectral radius is greater than one. The analogy ends there however. An affine dynamical system whose matrix has spectral radius one can exhibit several different behaviors: tendency toward infinity and periodicity as seen above, but also convergence to a fixed point, depending on the system and the initial condition.

Crumpet 34: Long Term Behavior

For an affine discrete dynamical system, $\mathbf{x}_{k+1} = M\mathbf{x}_k + \mathbf{b}$, where 1 is not an eigenvalue of M, the system has a unique fixed point, $\mathbf{x}^*$:

$$\begin{aligned} M\mathbf{x}^* + \mathbf{b} &= \mathbf{x}^* \\ M\mathbf{x}^* - \mathbf{x}^* &= -\mathbf{b} \\ (M - I)\mathbf{x}^* &= -\mathbf{b} \\ \mathbf{x}^* &= -(M - I)^{-1}\mathbf{b}. \end{aligned}$$

The inverse of $M - I$ exists because 1 is not an eigenvalue of M. Letting $\mathbf{y} = \mathbf{x} - \mathbf{x}^*$, which implies $\mathbf{x} = \mathbf{y} + \mathbf{x}^*$, and substituting into $\mathbf{x}_{k+1} = M\mathbf{x}_k + \mathbf{b}$:

$$\begin{aligned} \mathbf{y}_{k+1} + \mathbf{x}^* &= M(\mathbf{y}_k + \mathbf{x}^*) + \mathbf{b} \\ &= M\mathbf{y}_k + (M\mathbf{x}^* + \mathbf{b}) \\ &= M\mathbf{y}_k + \mathbf{x}^*. \end{aligned}$$

So $\mathbf{y}_{k+1} = M\mathbf{y}_k$. The analysis of this linear dynamical system fully informs the behavior of the affine system. Assuming M is diagonalizable, we set $\mathbf{y}_0 = \mathbf{x}_0 - \mathbf{x}^*$ (by substitution) and write $\mathbf{y}_0$ in terms of a basis of eigenvectors, $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then

$$\mathbf{y}_0 = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_n\mathbf{w}_n$$

for some scalars $c_1, c_2, \ldots, c_n$, and $\mathbf{y}_1 = M(c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_n\mathbf{w}_n) = c_1\lambda_1\mathbf{w}_1 + c_2\lambda_2\mathbf{w}_2 + \cdots + c_n\lambda_n\mathbf{w}_n$, $\mathbf{y}_2 = M(c_1\lambda_1\mathbf{w}_1 + c_2\lambda_2\mathbf{w}_2 + \cdots + c_n\lambda_n\mathbf{w}_n) = c_1\lambda_1^2\mathbf{w}_1 + c_2\lambda_2^2\mathbf{w}_2 + \cdots + c_n\lambda_n^2\mathbf{w}_n$, and so on:

$$\mathbf{y}_k = c_1\lambda_1^k\mathbf{w}_1 + c_2\lambda_2^k\mathbf{w}_2 + \cdots + c_n\lambda_n^k\mathbf{w}_n.$$

This solution is dominated by the nonzero term(s) with the eigenvalue(s) of greatest magnitude. If the dominant magnitude is less than one, $\mathbf{y}_k$ will tend toward zero and therefore $\mathbf{x}_k$ will tend toward $\mathbf{x}^*$, the fixed point. If the dominant magnitude is greater than one, $\mathbf{y}_k$ will tend toward infinity, in which case $\mathbf{x}_k$ will tend toward infinity.

Key Concepts

first order discrete dynamical system an equation of the form $\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k)$.

nonlinear discrete dynamical system a discrete dynamical system whose recurrence is a nonlinear transformation.

initial condition a value $\mathbf{v}$ for the first term of a dynamical system, usually given as $\mathbf{x}_0 = \mathbf{v}$.

recurrence the type of equation appearing in a discrete dynamical system.

recurrence relation a recurrence.

affine discrete dynamical system a dynamical system of the form $\mathbf{x}_{k+1} = M\mathbf{x}_k + \mathbf{b}$.

affine transformation a transformation $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ where $\mathbf{f}(\mathbf{x}) = M\mathbf{x} + \mathbf{b}$.

iteration the process of calculating the terms of the sequence determined by a discrete dynamical system.

iterates the terms of the sequence determined by a discrete dynamical system.

iterations iterates.

orbit the sequence determined by a discrete dynamical system—the solution of a dynamical system with initial condition.

long term behavior a quantitative or qualitative description of the tail end of the orbit of a dynamical system. The phrases "approaching the fixed point", "tending toward infinity", and "asymptotically periodic" are often used. Another possibility for nonlinear dynamical systems is "chaotic".

fixed point a solution of the equation $\mathbf{f}(\mathbf{x}) = \mathbf{x}$.

attractor the fixed point of a dynamical system whose solutions tend toward it.

repeller the fixed point of a dynamical system whose solutions tend away from it.

spectral radius the greatest magnitude of the eigenvalues of a matrix.

Exercises

1. Calculate the first four iterates of the dynamical system defined by $\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k)$ and $\mathbf{x}_0 = \mathbf{0}$.

   (a) $\mathbf{f}(\mathbf{x}) = \frac{1}{42}\begin{bmatrix} 16 & 26 \\ 65 & -23 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1/7 \\ 3/14 \end{bmatrix}$ [$\mathbb{S}$]-340

   (b) $\mathbf{f}(\mathbf{x}) = \frac{3}{2}\begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 7 \\ \frac{13}{2} \end{bmatrix}$

   (c) $\mathbf{f}(\mathbf{x}) = \frac{1}{2}\begin{bmatrix} -7 & -10 \\ 4 & 6 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 2 \\ -3 \end{bmatrix}$

   (d) $\mathbf{f}(\mathbf{x}) = \frac{1}{4}\begin{bmatrix} -6 & 5 \\ -4 & 3 \end{bmatrix}\mathbf{x} + \begin{bmatrix} \frac{5}{2} \\ \frac{5}{2} \end{bmatrix}$

   (e) $\mathbf{f}(\mathbf{x}) = \frac{1}{20}\begin{bmatrix} 40 & -36 \\ 63 & -56 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 3/5 \\ -7/10 \end{bmatrix}$ [$\mathbb{A}$]-365

   (f) $\mathbf{f}(\mathbf{x}) = \frac{1}{3}\begin{bmatrix} 3 & -4 \\ 6 & -7 \end{bmatrix}\mathbf{x} + \begin{bmatrix} -\frac{8}{3} \\ -\frac{14}{3} \end{bmatrix}$

   (g) $\mathbf{f}(\mathbf{x}) = \frac{1}{6}\begin{bmatrix} -3 & 3 \\ 3 & 5 \end{bmatrix}\mathbf{x} + \begin{bmatrix} -1 \\ 1/2 \end{bmatrix}$

   (h) $\mathbf{f}(\mathbf{x}) = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 2 \\ 12 \end{bmatrix}$

   (i) $\mathbf{f}(\mathbf{x}) = \begin{bmatrix} 11 & 10 & 0 \\ -8 & 5 & -12 \\ -8 & -4 & -3 \end{bmatrix}\mathbf{x} + \begin{bmatrix} -54 \\ -27 \\ 19 \end{bmatrix}$

   (j) $\mathbf{f}(\mathbf{x}) = \frac{1}{40}\begin{bmatrix} 155 & 64 & -55 \\ -192 & -76 & 72 \\ 147 & 64 & -47 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 3/5 \\ 3/8 \\ -7/10 \end{bmatrix}$ [$\mathbb{A}$]-365

   (k) $\mathbf{f}(\mathbf{x}) = \frac{1}{12}\begin{bmatrix} 4 & -16 & 8 \\ 2 & -22 & 9 \\ 4 & -20 & 6 \end{bmatrix}\mathbf{x} + \begin{bmatrix} \frac{26}{3} \\ \frac{85}{6} \\ \frac{13}{3} \end{bmatrix}$

   (l) $\mathbf{f}(\mathbf{x}) = \frac{1}{18}\begin{bmatrix} 82 & 57 & -29 \\ -16 & -3 & 41 \\ 8 & -3 & 65 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 5/6 \\ 7/3 \\ 1/18 \end{bmatrix}$

   (m) $\mathbf{f}(\mathbf{x}) = \frac{1}{28}\begin{bmatrix} -26 & 8 & 10 \\ -9 & -1 & 4 \\ -15 & 3 & 2 \end{bmatrix}\mathbf{x} + \begin{bmatrix} \frac{13}{14} \\ \frac{15}{14} \\ \frac{16}{7} \end{bmatrix}$

   (n) $\mathbf{f}(\mathbf{x}) = \frac{1}{28}\begin{bmatrix} 153 & 500 & 100 \\ -48 & -164 & -37 \\ 40 & 160 & 53 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1/7 \\ 2/7 \\ 3/7 \end{bmatrix}$ [$\mathbb{A}$]-365

   (o) $\mathbf{f}(\mathbf{x}) = -\frac{1}{6}\begin{bmatrix} -10 & -69 & 5 \\ 0 & 54 & 0 \\ 2 & 69 & -13 \end{bmatrix}\mathbf{x} + \begin{bmatrix} -\frac{569}{6} \\ 80 \\ \frac{583}{6} \end{bmatrix}$

2. Find the fixed point(s) of the dynamical system in question 1. [$\mathbb{S}$]-341 [$\mathbb{A}$]-366

3. Find the eigenvalues of the square matrix in question 1. [$\mathbb{A}$]-366

4. Based on the information from questions 1-3, does the dynamical system have an attractor? [$\mathbb{S}$]-341 [$\mathbb{A}$]-366

5. Describe the long term behavior of the dynamical system in question 1. Calculate more iterates if needed. [$\mathbb{S}$]-341 [$\mathbb{A}$]-366

6. Picturing an attractor, part 1. Let $M = \frac{1}{50}\begin{bmatrix} 24 & -8 \\ -3 & 26 \end{bmatrix}$. The fixed point of

$$\mathbf{x}_{k+1} = M\mathbf{x}_k + \begin{bmatrix} 4 \\ 6 \end{bmatrix};\ \mathbf{x}_0 = \mathbf{v}$$

   is $\begin{bmatrix} 4 \\ 12 \end{bmatrix}$ and the eigenpairs of M are $\frac{3}{5}, \begin{bmatrix} 4 \\ -3 \end{bmatrix}$ and $\frac{2}{5}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. [$\mathbb{A}$]-366

   (a) Verify that the spectral radius of M is less than 1 (and therefore the fixed point is an attractor).

   (b) On a set of axes, plot the fixed point with the eigenvectors emanating from it.

   (c) Pick several random points approximately 10 units away from the fixed point.

   (d) Calculate the first 4 iterations of the orbits of each point from part (c) and plot them on the same set of axes.

   (e) Connect each orbit with a single smooth arrow through its points in the order in which they occur.

7. Picturing an attractor, part 2. Let $M = \frac{\sqrt{2}}{3}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. The fixed point of

$$\mathbf{x}_{k+1} = M\mathbf{x}_k + \begin{bmatrix} 3 \\ 3 - 2\sqrt{2} \end{bmatrix};\ \mathbf{x}_0 = \mathbf{v}$$

   is $\begin{bmatrix} 3 \\ 3 \end{bmatrix}$ and the eigenvalues of M are $\frac{\sqrt{2}}{3}(1 \pm i)$. The eigenvectors are thus complex as well.

(a) Verify that the spectral radius of M is less than 1 (and therefore the fixed point is an attractor).

(b) On a set of axes, plot the fixed point.

(c) Pick several random points approximately 8 units away from the fixed point.

(d) Calculate the first 4 iterations of the orbits of each point from part (c) and plot them on the same set of axes.

(e) Connect the orbits by drawing a smooth arrow through them in the order in which they occur.

8. Picturing a repeller, part 1. Let $M = \frac{1}{15}\begin{bmatrix} 24 & -8 \\ -3 & 26 \end{bmatrix}$. The fixed point of

$$\mathbf{x}_{k+1} = M\mathbf{x}_k + \frac{1}{3}\begin{bmatrix} -32 \\ 29 \end{bmatrix};\ \mathbf{x}_0 = \mathbf{v}$$

is $\begin{bmatrix} 8 \\ -11 \end{bmatrix}$ and the eigenpairs of M are $2, \begin{bmatrix} 4 \\ -3 \end{bmatrix}$ and $\frac{4}{3}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

(a) Verify that both eigenvalues of M have magnitude greater than 1 (and therefore the fixed point is a repeller).

(b) On a set of axes, plot the fixed point with the eigenvectors emanating from it.

(c) Pick several random points about 1 unit from the fixed point.

(d) Calculate the first 4 iterations of the orbits of each point from part (c) and plot them on the same set of axes.

(e) Connect each orbit with a single smooth arrow through its points in the order in which they occur.

9. Picturing a repeller, part 2. Let $M = \frac{\sqrt{2}}{2}\begin{bmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{bmatrix}$. The fixed point of

$$\mathbf{x}_{k+1} = M\mathbf{x}_k + \begin{bmatrix} 3\sqrt{2} \\ 6 - 3\sqrt{6} \end{bmatrix};\ \mathbf{x}_0 = \mathbf{v}$$

is $\begin{bmatrix} 0 \\ 6 \end{bmatrix}$ and the eigenvalues of M are $\frac{\sqrt{2}}{2}(\sqrt{3} \pm i)$. The eigenvectors are thus complex as well.
[A]-366

(a) Verify that both eigenvalues have magnitude greater than 1 (and therefore the fixed point is a repeller).

(b) On a set of axes, plot the fixed point.

(c) Pick several random points approximately 1 unit away from the fixed point.

(d) Calculate the first 4 iterations of the orbits of each point from part (c) and plot them on the same set of axes.

(e) Connect the orbits by drawing a smooth arrow through them in the order in which they occur.

10. Let $M - I$ be an invertible matrix and suppose some fixed point of the affine dynamical system

$$\mathbf{x}_{k+1} = M\mathbf{x}_k + \mathbf{c};\ \mathbf{x}_0 = \mathbf{v}$$

is an attractor for any initial value $\mathbf{v}$. Argue that

(a) the fixed point is unique; and

(b) the fixed point of the dynamical system

$$\mathbf{z}_{k+1} = M^{-1}(\mathbf{z}_k - \mathbf{c});\ \mathbf{z}_0 = \mathbf{v}$$

is unique and is a repeller.

11. Let $M - I$ be an invertible matrix and suppose some fixed point of the affine dynamical system

$$\mathbf{x}_{k+1} = M\mathbf{x}_k + \mathbf{c};\ \mathbf{x}_0 = \mathbf{v}$$

is a repeller for any initial value $\mathbf{v}$. Argue that

(a) the fixed point is unique; and

(b) the fixed point of the dynamical system

$$\mathbf{z}_{k+1} = M^{-1}(\mathbf{z}_k - \mathbf{c});\ \mathbf{z}_0 = \mathbf{v}$$

is unique and is an attractor.

12. Picturing a saddle point. Let $M = \frac{1}{6}\begin{bmatrix} 7 & -3 \\ -1 & 5 \end{bmatrix}$. The fixed point of

$$\mathbf{x}_{k+1} = M\mathbf{x}_k + \frac{1}{6}\begin{bmatrix} -13 \\ 1 \end{bmatrix};\ \mathbf{x}_0 = \mathbf{v}$$

is $\begin{bmatrix} 4 \\ -3 \end{bmatrix}$ and the eigenpairs of M are $\frac{4}{3}, \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ and $\frac{2}{3}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

(a) Verify that one eigenvalue has magnitude greater than 1 while the other has magnitude less than 1 (and therefore the fixed point is neither an attractor nor repeller).

(b) On a set of axes, plot the fixed point with the eigenvectors emanating from it.

(c) Pick several random points closer to the line defined by the eigenvector whose corresponding eigenvalue has magnitude greater than 1 than they are to the line defined by the other eigenvector.

(d) Calculate the first 4 iterations of the orbits of each point from part (c) and plot them on the same set of axes.

(e) Connect each orbit with a single smooth arrow through its points in the order in which they occur.

[7] See Discrete Dynamical Systems: With Applications in Biology.

13. Host Parasite Interaction. Let H be the population of a host prone to parasite population P and suppose the populations evolve according to the discrete dynamical system

$$\begin{aligned} H_{n+1} &= \gamma H_n e^{-aP_n} \\ P_{n+1} &= H_n\left(1 - e^{-aP_n}\right) \end{aligned} \tag{7.4.10}$$

for some positive values of the parameters γ and a.[7]

(a) Find the fixed point, $\begin{bmatrix} H^* \\ P^* \end{bmatrix}$, of (7.4.10).

(b) For $0 < \gamma < 1$, argue that the fixed point of (7.4.10) has a negative coordinate (and therefore is an unattainable state for the physical system—populations cannot be negative).

(c) For $\gamma = 1$, argue that the parasite population is zero according to (7.4.10).

(d) The linear dynamical system

$$\begin{bmatrix} \hat{H} \\ \hat{P} \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & \frac{\gamma \ln\gamma}{1-\gamma} \\ \frac{\gamma-1}{\gamma} & \frac{\ln\gamma}{\gamma-1} \end{bmatrix} \begin{bmatrix} \hat{H} \\ \hat{P} \end{bmatrix}_{k} \tag{7.4.11}$$

where $\begin{bmatrix} \hat{H} \\ \hat{P} \end{bmatrix} = \begin{bmatrix} H - H^* \\ P - P^* \end{bmatrix}$, called a linearization of (7.4.10), is an excellent approximation of (7.4.10) near its fixed point. For $\gamma > 1$, argue that the fixed point of (7.4.11) is a repeller (and therefore is an unstable state for the physical system—populations will tend away from the fixed point). This is enough to show that the same happens for (7.4.10).

(e) Is this a good model for host/parasite populations that can live in equilibrium (constant populations each)? Explain.

Answers

brownie iterates Given that $\mathbf{x}_0 = \begin{bmatrix} 72 \\ 66 \end{bmatrix}$, accurate to three decimal places,

$$\begin{aligned}
\mathbf{x}_1 &= \frac{1}{146}\begin{bmatrix} 146 & 0 \\ 1 & 145 \end{bmatrix}\begin{bmatrix} 72 \\ 66 \end{bmatrix} + \begin{bmatrix} 21 \\ 0 \end{bmatrix} = \begin{bmatrix} 93 \\ 66.041 \end{bmatrix} \\
\mathbf{x}_2 &= \frac{1}{146}\begin{bmatrix} 146 & 0 \\ 1 & 145 \end{bmatrix}\begin{bmatrix} 93 \\ 66.041 \end{bmatrix} + \begin{bmatrix} 21 \\ 0 \end{bmatrix} = \begin{bmatrix} 114 \\ 66.226 \end{bmatrix} \\
\mathbf{x}_3 &= \frac{1}{146}\begin{bmatrix} 146 & 0 \\ 1 & 145 \end{bmatrix}\begin{bmatrix} 114 \\ 66.226 \end{bmatrix} + \begin{bmatrix} 21 \\ 0 \end{bmatrix} = \begin{bmatrix} 135 \\ 66.553 \end{bmatrix} \\
\mathbf{x}_4 &= \frac{1}{146}\begin{bmatrix} 146 & 0 \\ 1 & 145 \end{bmatrix}\begin{bmatrix} 135 \\ 66.553 \end{bmatrix} + \begin{bmatrix} 21 \\ 0 \end{bmatrix} = \begin{bmatrix} 156 \\ 67.022 \end{bmatrix}
\end{aligned}$$

so the first five iterates of the orbit are

$$\begin{bmatrix} 72 \\ 66 \end{bmatrix}, \begin{bmatrix} 93 \\ 66.041 \end{bmatrix}, \begin{bmatrix} 114 \\ 66.226 \end{bmatrix}, \begin{bmatrix} 135 \\ 66.553 \end{bmatrix}, \begin{bmatrix} 156 \\ 67.022 \end{bmatrix}.$$

last iteration example Sample SageMath code that can be copied and pasted into a SageCell:

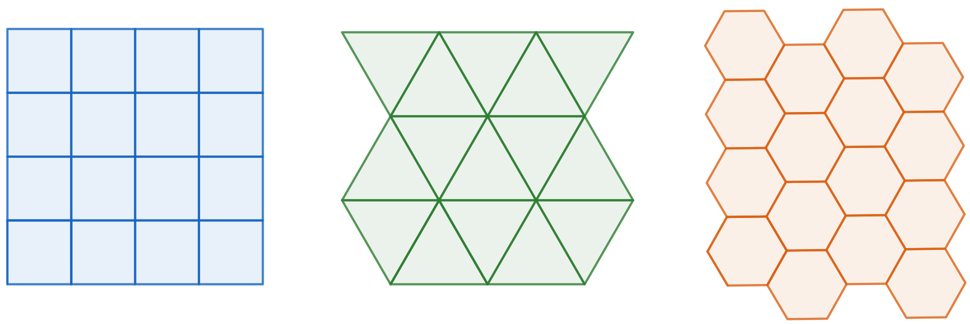
```
M=1/4*matrix(2,2,[sqrt(6.0)+sqrt(2.0), sqrt(2.0)-sqrt(6.0),
                  sqrt(6.0)-sqrt(2.0), sqrt(6.0)+sqrt(2.0)])
b=vector([-2.0,2.0])
x0=vector([1,1])
print("0 :",x0)
for i in range(1,5):
    x0=M*x0+b
    print(i,":",x0)
```

7.5 Rep-tiles [6.3]

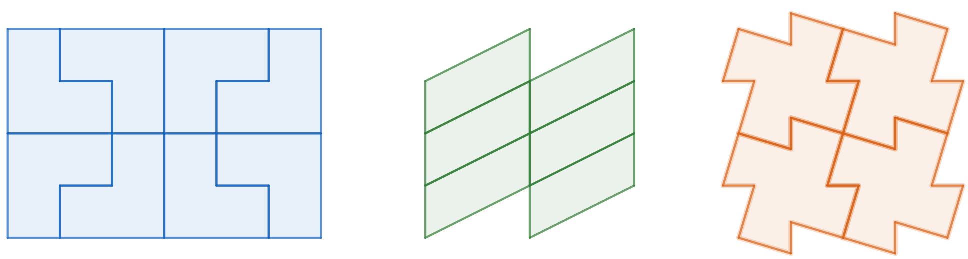
The floors of kitchens, bathrooms, museums, and other spaces are tiled more often than not. Flat ceramic or natural stone tiles are placed together in a nonoverlapping way, covering the whole floor. It is very common to see square tiles laid out in a grid, for example. Squares are easy to fit together this way and the pattern can be extended to cover any amount of space.

The plane $\mathbb{R}^2$ can be imagined as a floor without boundaries. Covering it with tiles requires extending those tiles endlessly in every direction. The most familiar example is a boundless rectangular grid. Imagine rectangular graph paper extended forever in every direction. It may not be the most attractive covering of the plane, but it does the job.

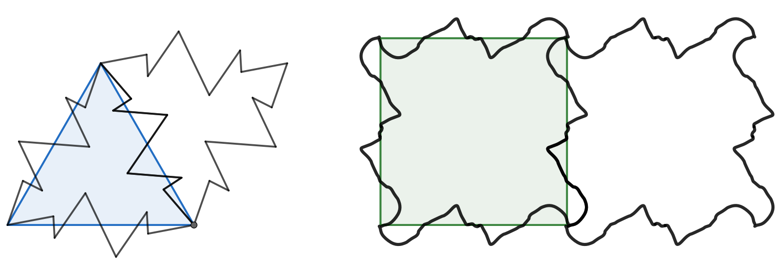
Any set of shapes covering the plane without overlapping is called a tessellation, and the shapes are said to tessellate the plane. Like squares, equilateral triangles and regular hexagons can be fitted together to tessellate the plane in a simple pattern. In fact these three shapes form the bases for the only three so-called regular tessellations, portions of which are shown here. Only your imagination can extend the patterns indefinitely.



Polygons that are not regular tessellate the plane just as well. Parallelograms, hexagons, dodecagons, convex and concave, can all be fitted to tessellate the plane. Portions of a small sample are shown here.



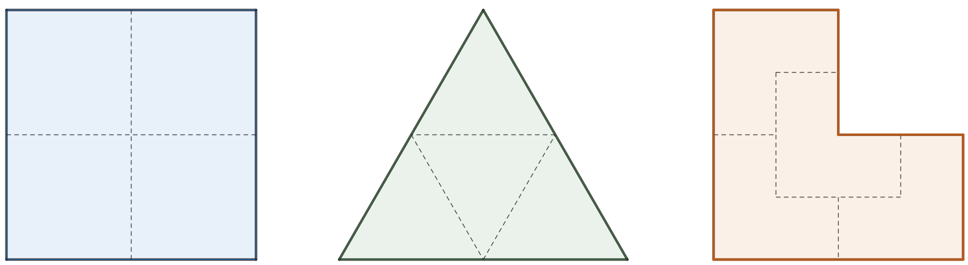
M.C. Escher famously made tessellation an artform. Tilings with irregularly shaped birds, fish, human figures, and other natural shapes appear in many of his most famous creations. See figure 7.5.1, for example. One way to create Escher-esque tessellations is to start with a regular tile and modify its perimeter in a symmetric way. The third tessellation of the diagram above is created from squares where each side is replaced by a zig-zag, for example. Among the infinite possibilities for tiles created this way are the two shown here.



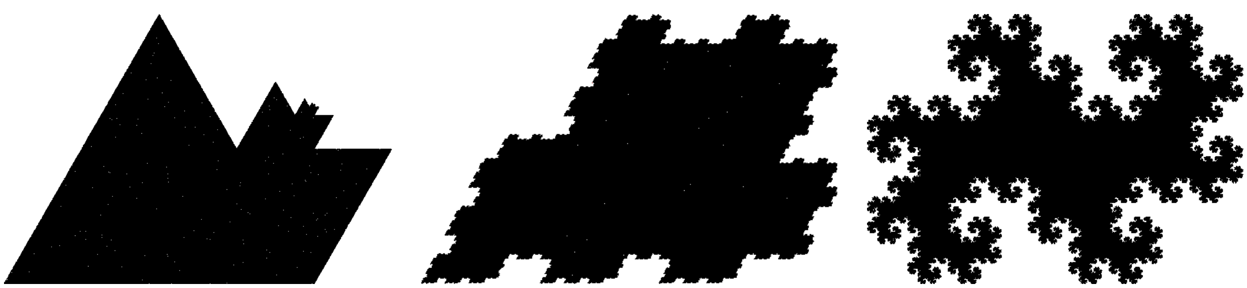
Each edge of the regular tile is replaced by a curve with 180 degree rotational symmetry about the midpoint of the original edge. As long as the replacements do not intersect one another, the resulting shape is a tile. That is, multiple copies can be fitted together to cover, or tessellate, the plane.

Figure 7.5.1: M.C. Escher's System X(e)

Certain tiles (shapes that tessellate the plane) are actually doubly tiling. Not only can they be fitted together to tile the plane—they can also be fitted together to tile larger copies of themselves! These shapes are called **rep-tiles**, short for self-replicating tiles. Again, the square provides an immediate example. Four congruent squares fitted together at a corner, sides parallel to one another form a square with side length twice the original (and four times the area). Equilateral triangles, and in fact all triangles, are rep-tiles. Four congruent copies can be pieced together, three in the same orientation and the fourth rotated 180 degrees, to form a larger copy. Regular hexagons are not rep-tiles as no finite number of copies of a hexagon can be fitted together (as tiles, without overlap) to form a hexagon. However, there are many non-regular rep-tilian hexagons. For example, the hexagon formed by gluing three squares together in an ell is a rep-tile. The following diagram demonstrates the self-replication of a square, a regular triangle, and this hexagonal ell shape. Four copies of each shape are fitted together to form an enlarged replica. Since we have already seen that these shapes tessellate the plane, they are indeed rep-tiles.



Rep-tiles come in much more fanciful shapes, however. Take these three, for example.



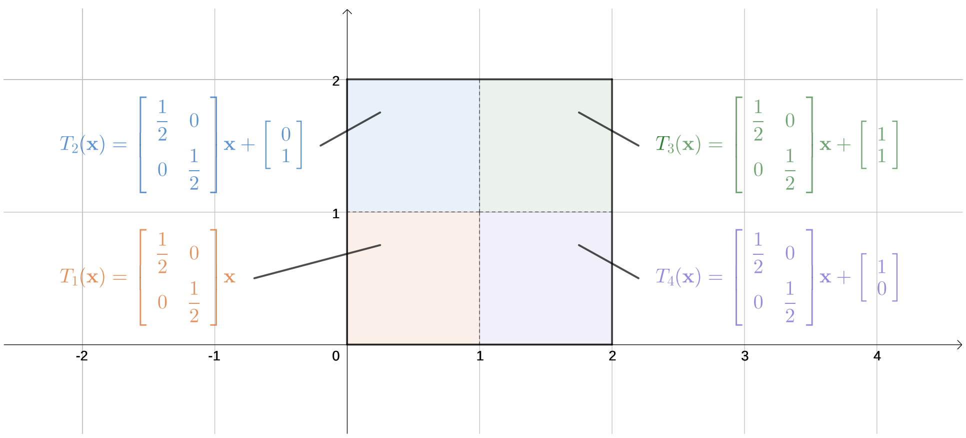
Appearing rom left to right are the carpenter's plane of Golomb[10], a fractile of Bandt[2], and a twindragon. Seeing that these shapes are in fact rep-tiles is nontrivial. Pictures showing them tiling replicas of themselves and tessellating a portion of the plane would make adequate demonstrations, but would give no insight into their origin or how to imagine others, or even how to define the shapes themselves. For that, we rely on linear algebra.

After building a replica of a rep-tile (from similar copies of the rep-tile itself), one can switch perspecives and look at the completed figure as a dissection of the larger rep-tile. From this viewpoint, rep-tiles are plane figures that tessellate the plane and can be dissected into finitely many similar copies of themselves. All rep-tiles can be seen from this vantage. Refer back to the diagram of the square, the equlateral triangle and the hexagonal ell being fitted together to self-replicate with a different lens. The square is shown dissected into four smaller squares. The equilateral triangle is shown dissected into four smaller equilateral triangles, and the ell shaped hexagon is likewise divided into four smaller copies of itself.

Crumpet 35: Solomon Golomb

Solomon Golomb is credited with coining the term *rep-tile*, but his original paper[10] only uses the term “rep-*k*”, short for *replicating of order k*. To quote Golomb, a plane figure is called rep-*k* if “it can be dissected into *k* ‘replicas’, each congruent to the others and similar to the original”. Curiously, Martin Gardner[7] credits Golomb with laying the foundation for the study of rep-tiles and inventing the term in a series of private papers, all in his article appearing more than a year earlier than Golomb’s!

By imposing a set of axes on any of these figures, rigorous mathematical descriptions of the figures become available. Any placement of the axes will do. Only a frame of reference is needed. For example, suppose we arrange for opposite corners of the square to coincide with $(0,0)$ and $(2,2)$. The following diagram shows the four squares of its dissection, and for each of these squares, an affine transformation mapping the whole square to the part.



In words, the 2×2 matrices scale shapes (and more to the point, the square) by a factor of $\frac{1}{2}$ in both the horizontal and vertical directions. The addition of 2×1 vectors provide translations. T_4 can thus be described as scaling by $\frac{1}{2}$ horizontally and vertically followed by translation 1 unit horizontally. $T_4\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $T_4\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, for example. Again, the critical point is that the image of the 2 by 2 square under T_4 is the purple square: contracting the 2 by 2 square by a factor of $\frac{1}{2}$ and then translating the contracted copy right 1 unit lands the image of the larger square (well, squarely) on top of the purple square—the bottom right square of the dissection. Letting S be the 2 by 2 square with opposite corners at $(0,0)$ and $(2,2)$, we thereby have $T_4(S) =$ the purple square. Similarly, $T_1(S) =$ the orange square (the bottom left square of the dissection); $T_2(S) =$ the blue square (the top left square of the dissection); and $T_3(S) =$ the green square (the top right square of the dissection).

The union of the four images, $T_1(S)$, $T_2(S)$, $T_3(S)$, and $T_4(S)$, is the original square. In the form of an equation,

$$T_1(S) \cup T_2(S) \cup T_3(S) \cup T_4(S) = S. \tag{7.5.1}$$

A theorem of Hutchinson[14] asserts that S is the only compact set that satisfies (7.5.1). In other words, the square S is determined by the transformations T_1, T_2, T_3, T_4 via equation (7.5.1). This way, these four transformations provide a precise description, or definition, of the square with opposite corners at $(0,0)$ and $(2,2)$. Incidentally, each transformation T_k is a **similitude**—a **rigid transformation** (rotation, reflection, translation, or composition thereof) composed with **dilation** (scaling by the same scale factor in all directions). Similitudes preserve shape but not necessarily size, exactly the type of transformation needed to map a shape onto one of the (similar) parts of its dissection.

Let $\mathscr{C} = \{C_1, C_2, \ldots, C_p\}$ be a set of similitudes in $\mathbb{R}^n$ with scale factors less than one, and define

$$\mathcal{H}_{\mathscr{C}}(A) = C_1(A) \cup C_2(A) \cup \cdots \cup C_p(A) \tag{7.5.2}$$

for any subset A of $\mathbb{R}^n$. **Hutchinson's theorem** concludes that there is exactly one compact set K in $\mathbb{R}^n$ such that $\mathcal{H}_{\mathscr{C}}(K) = K$. Moreover,

$$\lim_{k\to\infty} \mathcal{H}_{\mathscr{C}}^{\circ k}(A) = K \tag{7.5.3}$$

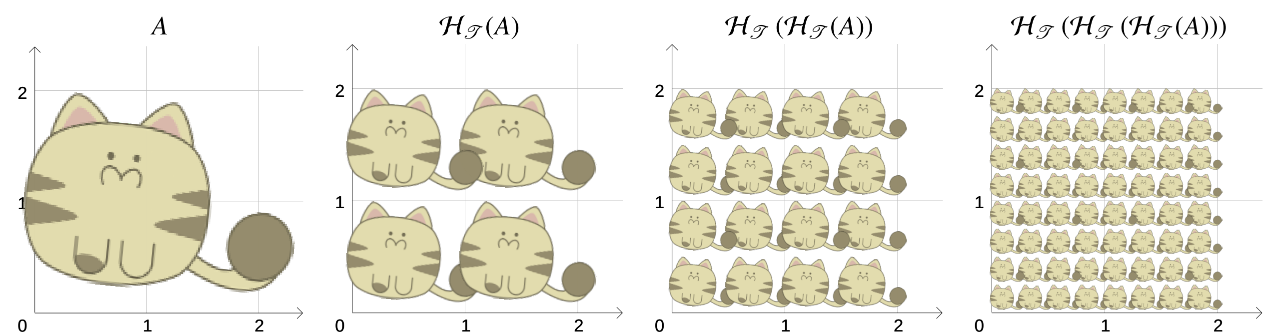
for any compact set A. Not only does the theorem assert the existence and uniqueness of the set K, it gives a way to construct it from the similitudes.

Crumpet 36: Hutchinson

The original theorem of Hutchinson and its proof lie along the fence between real analysis and topology. *Let $X = (X, d)$ be a complete metric space and $\mathscr{S} = \{S_1, \ldots, S_N\}$ be a finite set of contraction maps on X. Then there exists a unique closed bounded set K such that $K = \bigcup_{i=1}^N S_i K$. Furthermore, K is compact and is the closure of the set of fixed points $s_{i_1\cdots i_p}$ of finite compositions $S_{i_1} \circ \ldots \circ S_{i_p}$ of members of $\mathscr{S}$.*

For arbitary $A \subset X$ let $\mathscr{S}(A) = \bigcup_{i=1}^N S_i A$, $\mathscr{S}^P(A) = \mathscr{S}(\mathscr{S}^{P-1}(A))$. Then for closed bounded A, $\mathscr{S}^P(A) \to K$ in the Hausdorff metric.

Applying this theorem to the set $\mathscr{T} = \{T_1, T_2, T_3, T_4\}$, we do not have to know anything about the origin of the transformations T_k. All the work of dissecting the square, placing it in the plane, and deriving the transformations in $\mathscr{T}$ can be forgotten. All we need is a compact set A (and a lot of patience!) to recover the square. It is the limit of the sequence $\mathcal{H}_{\mathscr{T}}(A), \mathcal{H}_{\mathscr{T}}(\mathcal{H}_{\mathscr{T}}(A)), \mathcal{H}_{\mathscr{T}}(\mathcal{H}_{\mathscr{T}}(\mathcal{H}_{\mathscr{T}}(A))), \ldots$, the iteration of $\mathcal{H}_{\mathscr{T}}$ on *any* compact set A. The first few terms of this sequence are shown below, where A takes the shape of a kitty[8].



The following set of similitudes defines the hexagonal ell within the square with opposite vertices $(0, 0)$ and $(2, 2)$.

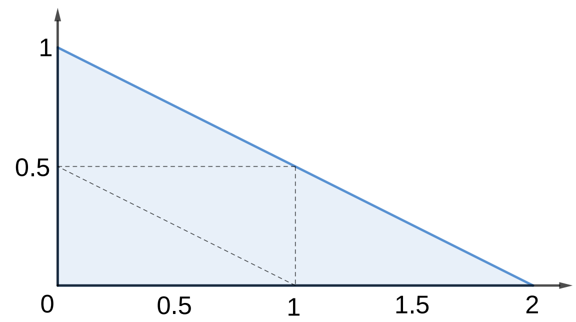
$$L_1(\mathbf{x}) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{x} \qquad L_2(\mathbf{x}) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{x} + \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

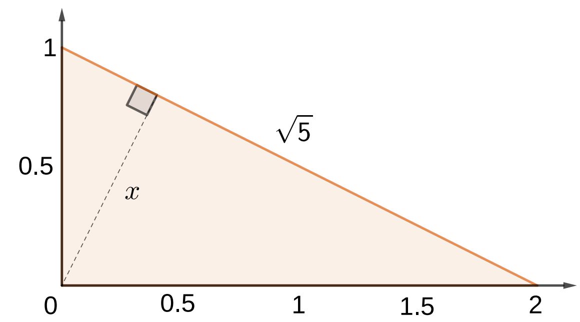
$$L_3(\mathbf{x}) = \begin{bmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 2 \\ 0 \end{bmatrix} \qquad L_4(\mathbf{x}) = \begin{bmatrix} 0 & \frac{1}{2} \\ -\frac{1}{2} & 0 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

Can you generate the ell shaped rep-tile by applying (7.5.3) to $\mathscr{C} = \{L_1, L_2, L_3, L_4\}$ (and some set A of your own creation)? Answer on page 277.

The following diagram illustrates three things. One, the dissection of a rep-tile is not unique (two different dissections are shown for the same right triangle). Two, the number of parts in a dissection of a rep-tile is not always four. Three, the parts of a dissection need not be congruent to one another (they must only be similar to the whole).

[8]Kitty image downloaded from https://openclipart.org/detail/292277/cute-cat.





Can you find the similitudes associated with these dissections (note there will be four similitudes associated with the first dissection and two similitudes associated with the second)? Answer on page 277.

In the second dissection, the two scale factors are $\frac{1}{\sqrt{5}}$ and $\frac{2}{\sqrt{5}}$. Not coincidentally $\left(\frac{1}{\sqrt{5}}\right)^2 + \left(\frac{2}{\sqrt{5}}\right)^2 = 1$. In general, if $\mathscr{C} = \{T_1, T_2, \ldots, T_p\}$ is the set of similitudes that determine a rep-tile R, then

$$T_1(R) \cup T_2(R) \cup \cdots \cup T_p(R) = R$$

and the images $T_k(R)$ pariwise have no overlapping area, so

$$\text{area}\,(T_1(R)) + \text{area}\,(T_2(R)) + \cdots + \text{area}\left(T_p(R)\right) = \text{area}(R). \tag{7.5.4}$$

Now, we know from section 6.3 that $\text{area}\,(T_k(R)) = |\det L_k| \cdot \text{area}(R)$ where L_k is the matrix of the linear part of similitude T_k. We also know from section 3.7 that the determinant of a product is the product of the determinants and from section 3.5 that multiplying a 2×2 matrix by scalar c multiplies its determinant by c^2. Finally, combined with the facts that the determinants of reflections and rotations are -1 and 1 respectively, the determinant of any matrix of a similitude is s^2 where s is its scale factor. Applying this information to equation (7.5.4), we get

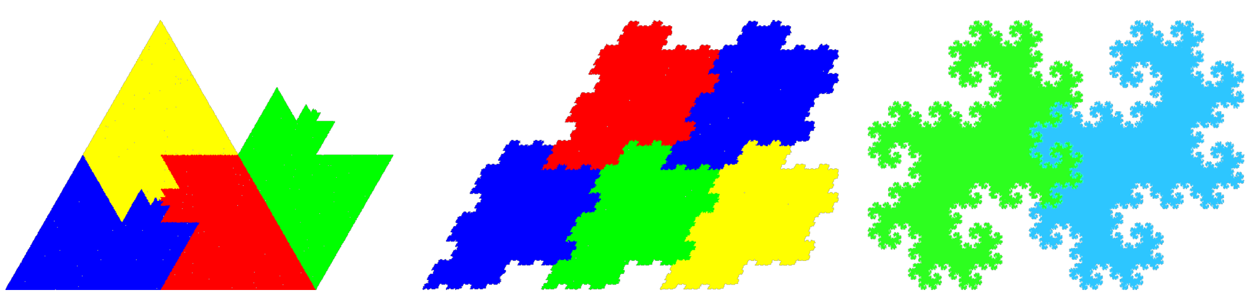
$$s_1^2\text{area}(R) + s_2^2\text{area}(R) + \cdots + s_p^2\text{area}(R) = \text{area}(R)$$

where s_k is the scale factor of similitude T_k. Hence

$$s_1^2 + s_2^2 + \cdots + s_p^2 = 1. \tag{7.5.5}$$

The square, the hexagonal ell, and the triangle were all dissected into four parts, each of which was a $\frac{1}{2}$ scale replica of the whole. By equation (7.5.5) it must be that $\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = 1$, an equality that is not hard to verify. As a matter of vocabulary, the set of similitudes associated with a rep-tile is an **iterated function system**, or **IFS**. Hence, if $s_1, s_2, \ldots, s_p$ are the scale factors of the similitudes of the IFS of a rep-tile, then $s_1^2 + s_2^2 + \cdots + s_p^2 = 1$.

Returning to the carpenter's plane of Golomb, the fractile of Bandt, and the twindragon, shown below are dissections.



Imposing a set of axes on any one of the dissections allows developing the similitudes mapping the shape to its parts. Much like a center and radius define a circle or two points define a line, the collection of these similitudes defines the rep-tile.

Key Concepts

rep-tile a plane figure that tessellates the plane and can be dissected into finitely many similar copies. Equivalently, a plane figure that tiles the plane and tiles an enlarged replica of itself.

Hutchinson's theorem (a special case) Let $\mathscr{C} = \{C_1, C_2, \ldots, C_n\}$ be a set of similitudes in $\mathbb{R}^2$ with scale factors less than one, and define

$$\mathcal{H}_{\mathscr{C}}(A) = C_1(A) \cup C_2(A) \cup \cdots \cup C_n(A) \tag{7.5.6}$$

for any subset A of $\mathbb{R}^2$. Then there is exactly one compact set K in $\mathbb{R}^2$ such that $\mathcal{H}_{\mathscr{C}}(K) = K$. Moreover,

$$\lim_{k\to\infty} \mathcal{H}_{\mathscr{C}}^{\circ k}(A) = K \tag{7.5.7}$$

for any compact set A.

similitude a rigid transformation composed with a dilation. Similitudes preserve shape but not necessarily size

rigid transformation a rotation, reflection, or translation.

dilation a map of the form $T(\mathbf{x}) = r\mathbf{x}$ for some real number $r \geq 0$—scaling by the same factor in all directions.

compact set a subset S of $\mathbb{R}^n$ is compact if it is closed and bounded.

closed set a subset S of $\mathbb{R}^n$ is closed if the limit of every convergent sequence of points in S is also in S. Alternatively, S is closed if it contains all of its limit points.

bounded set a subset S of $\mathbb{R}^n$ is bounded if there exists a real number M such that $S \subseteq \{\mathbf{x} \text{ in } \mathbb{R}^n : ||\mathbf{x}|| < M\}$. Alternatively, S is bounded if it is contained within some ball centered at the origin.

iterated function system a set of contraction mappings.

contraction mapping a map $T : \mathbb{R}^n \to \mathbb{R}^n$ is a contraction (mapping) if for every distinct pair of points $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ there exists a real number $s < 1$ such that

$$\frac{d(T(\mathbf{x}), T(\mathbf{y}))}{d(\mathbf{x}, \mathbf{y})} \leq s.$$

A contraction mapping scales down the distance between every pair of distinct points. A similitude with scale factor less than one is a contraction mapping.

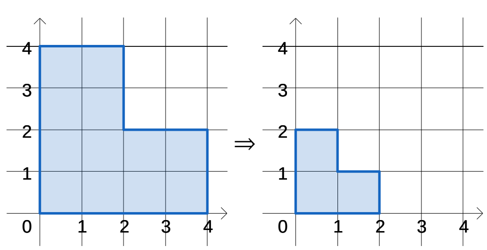
IFS iterated function system.

scale factors of the IFS of a rep-tile if $s_1, s_2, \ldots, s_p$ are the scale factors of the similitudes of the IFS of a rep-tile, then $s_1^2 + s_2^2 + \cdots + s_p^2 = 1$.

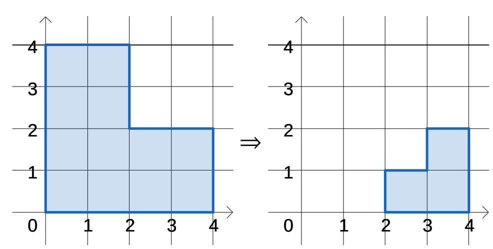
Exercises

1. Determine an affine transformation that maps the large figure to the similar (smaller) figure.

(a)

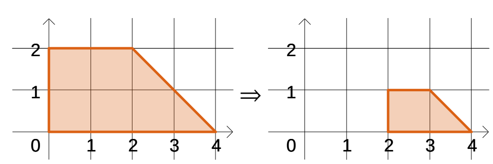


(b)

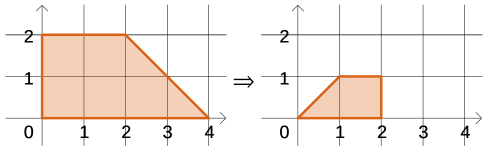


[𝕊]-341

(c)

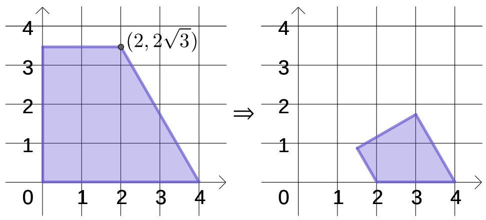


(d)

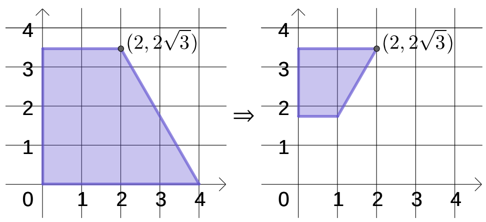


[A]-366

(e)

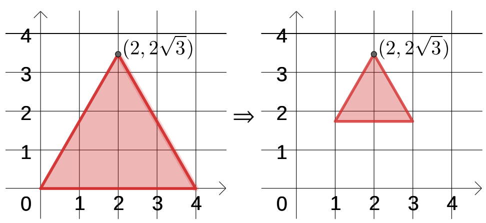


(f)

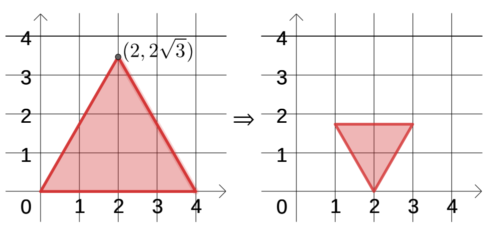


[A]-366

(g)



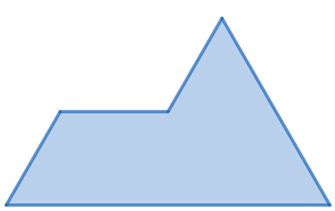
(h)



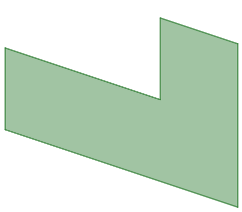
[A]-366

2. Build a larger copy of the figure from similar copies of itself, thereby showing that it has one feature of a rep-tile. Appropriate sizes and number of copies can be found on page 278.

(a)

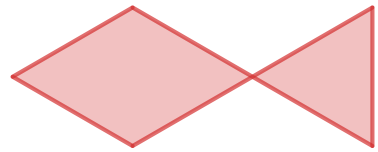


(b)

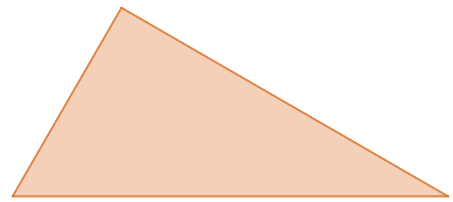


[A]-366

(c)

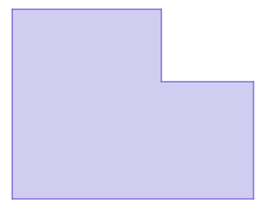


(d)

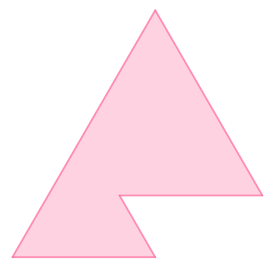


[A]-366

(e)



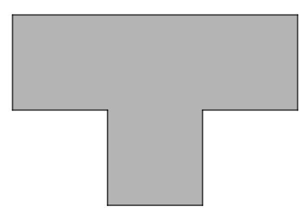
(f)



[A]-366

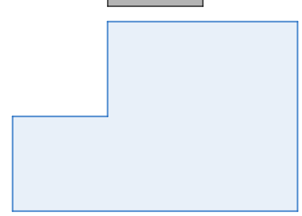
(g)

(h)



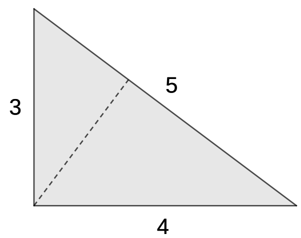
[A]-366

(i)



(j)

3. Show that the shape in question 2 tessellates the plane. This completes a demonstration that the shape is a rep-tile. [𝕊]-342 [𝔸]-367
4. The 3-4-5 right triangle is a rep-tile that can be dissected into two parts as shown.



   (a) Use similar triangles to calculate the scale factors of the two similitudes, s_1 and s_2, of its IFS.

   (b) Verify that $s_1^2 + s_2^2 = 1$.

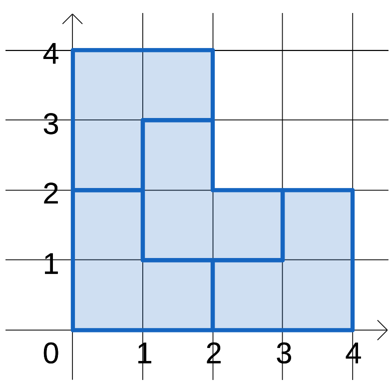
5. Find a dissection of the 30-60-90 triangle into three congruent parts, each similar to the whole, showing that the triangle is a rep-tile. Equivalently, build a 30-60-90 triangle[9] out of three congruent 30-60-90 triangles.
6. The right triangle with side lengths $1, 2$, and $\sqrt{5}$ can be dissected into five congruent parts, each similar to the whole, two different ways. Find one of them. [𝔸]-367
7. What are the scale factors of the IFS for the triangle in question

   (a) 5

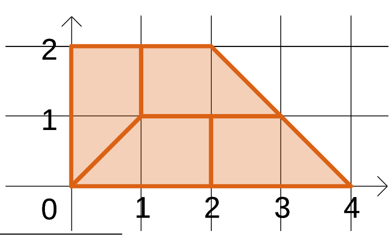
   (b) 6 [𝔸]-367

8. A rectangle can be dissected into three congruent parts, making it a rep-tile. Equivalently, three congruent copies of this rectangle can be fitted together to form a (larger) similar copy. What is the ratio of its side lengths?
9. Find an IFS of the rep-tile suggested by the dissection. Impose your own set of axes where not supplied.

   (a)

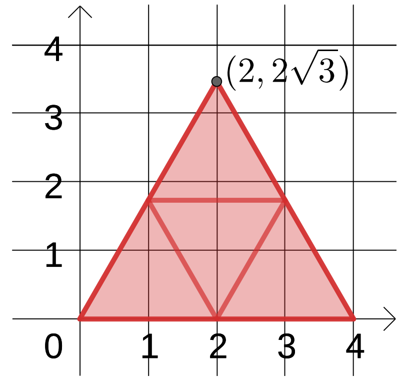


   (b)

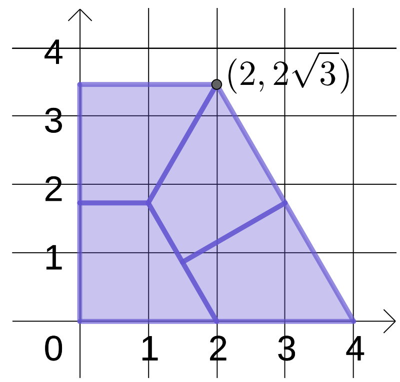


[𝔸]-367

   (c)

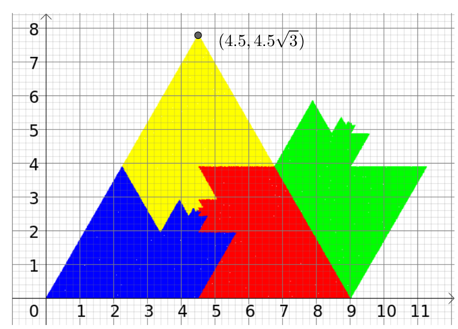


   (d)

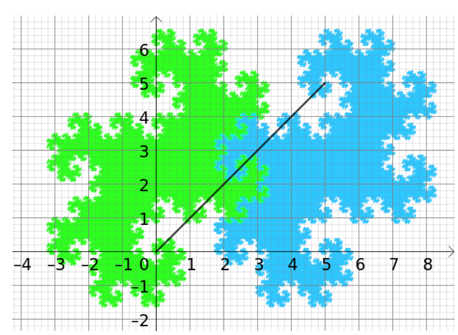


[𝕊]-342

   (e)



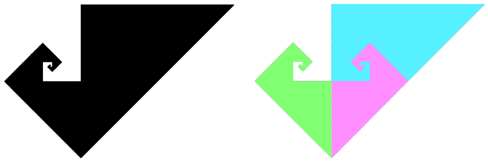
   (f) The images of the line segment under the similtudes of the IFS are the line segments from $(0,0)$ to $(0,5)$ and from $(5,0)$ to $(5,5)$.



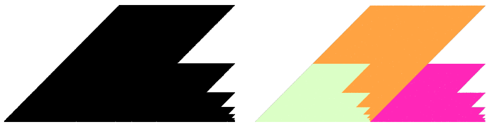
[𝔸]-367

[9] A 30-60-90 triangle is one whose interior angles measure 30, 60, and 90 degrees.

(g)

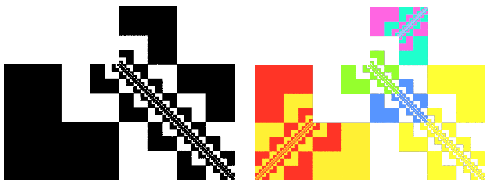


(h)

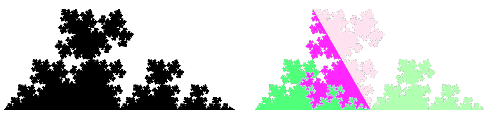


[𝔸]-367

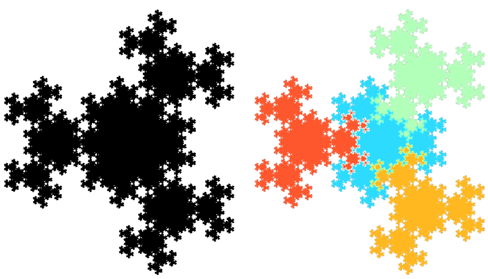
(i)



(j)

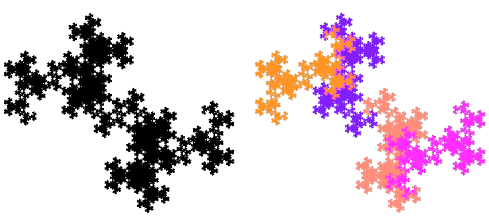


(k)



[𝔸]-367

(l)



10. Check your answers for question 9 with the rep-tile designer.[10] It will be helpful to have your similitudes written in terms of the designer's format. Each similitude should be expressed as a composition of

    (i) a reflection (across the x-axis, y-axis or neither)

    (ii) a scaling (scale factor)

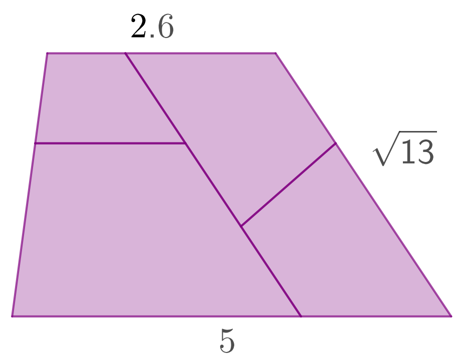
    (iii) a rotation (in degrees about the origin)

    (iv) a horizontal translation

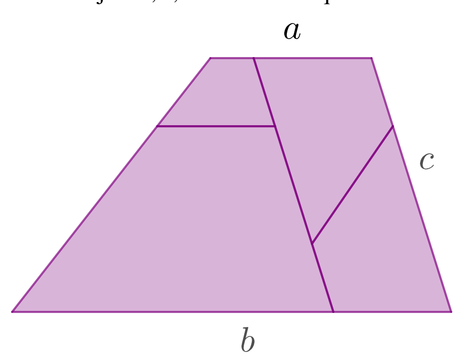
    (v) a vertical translation

    in that order. [𝕊]-343 [𝔸]-367

11. Find the three scale factors of the IFS suggested by the dissection. [𝕊]-344



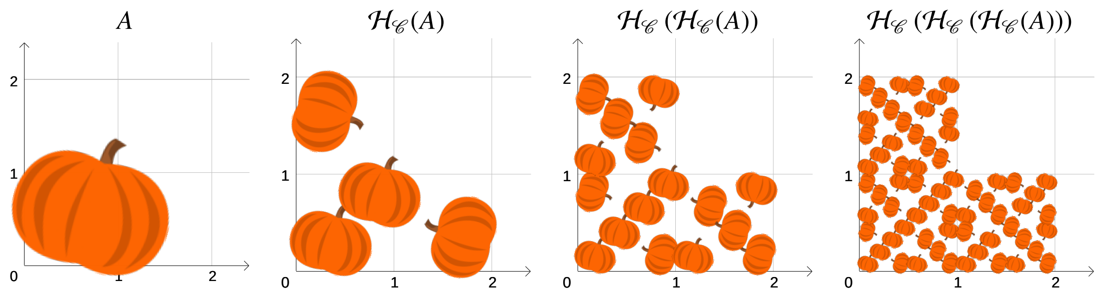
12. Find an expression for c in terms of a and b. HINT: Write down five equations involving the three scale factors of the IFS suggested by the dissection. Four of them can be used to eliminate the scale factors, leaving a single equation with just a, b, c. Solve this equation for c.



[10] https://lqbrin.github.io/tea-time-linear/rep-tile-designer.html

Answers

generating the L-shape Letting A take the shape of a pumpkin[11], the L-shape appears rather plainly after only three iterations:



dissecting the triangle For the first dissection, the four transformations mapping the whole triangle to the four parts are, in words,

1. scale by a factor of $\frac{1}{2}$,
2. scale by a factor of $\frac{1}{2}$ and then translate 1 unit right,
3. scale by a factor of $\frac{1}{2}$ and then translate $\frac{1}{2}$ unit up, and
4. scale by a factor of $\frac{1}{2}$, rotate (about the origin) by 180°, and then translate $\frac{1}{2}$ unit up and one unit right.

As affine transformations, the mappings are

$$\mathbf{x} \mapsto \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \mathbf{x}; \qquad \mathbf{x} \mapsto \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \mathbf{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix};$$
$$\mathbf{x} \mapsto \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix}; \qquad \mathbf{x} \mapsto \begin{bmatrix} -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix}.$$

For the second dissection, remember all the triangles are similar, so corresponding parts are in proportion. In particular, the smallest triangle is a $\frac{1}{\sqrt{5}}$ scaled version of the whole and the remaining part is a $\frac{2}{\sqrt{5}}$ scaled version. Getting a little ahead of ourselves, transformations mapping the whole triangle to the two parts are, in words,

1. scale by a factor of $\frac{1}{\sqrt{5}}$, reflect about the y-axis, rotate by angle β (counterclockwise about the origin), then translate along line segment x; and
2. scale by a factor of $\frac{2}{\sqrt{5}}$, reflect about the x-axis, rotate by angle $-\theta$ (clockwise about the origin), then translate along line segment x.

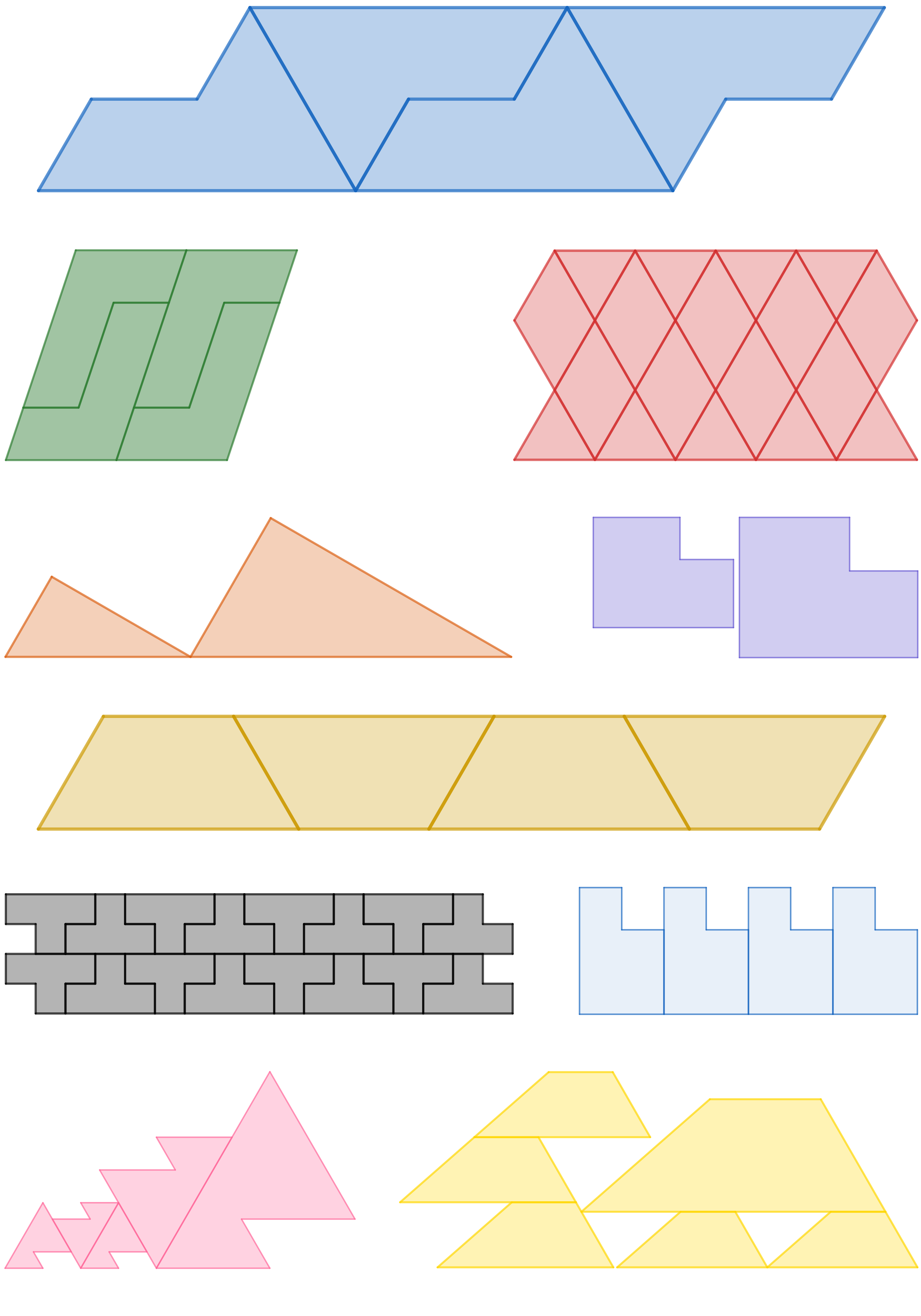
To quantify the rotations and translations, we need to calculate x and the sines and cosines of β and θ. Using the Pythagorean theorem, $1^2 = x^2 + \left(\frac{x}{2}\right)^2$ so $x = \frac{2}{\sqrt{5}}$, and the coordinates of P are $(x\cos\beta, x\sin\beta)$. But $\cos\beta = \frac{1}{\sqrt{5}}$ and $\sin\beta = \frac{2}{\sqrt{5}}$. Finally $\cos\theta = \frac{2}{\sqrt{5}}$ and $\sin\theta = \frac{1}{\sqrt{5}}$, so the mappings are

$$\mathbf{x} \mapsto \frac{1}{\sqrt{5}} \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x} + \begin{bmatrix} \frac{2}{5} \\ \frac{4}{5} \end{bmatrix}; \; \mathbf{x} \mapsto \frac{2}{\sqrt{5}} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{x} + \begin{bmatrix} \frac{2}{5} \\ \frac{4}{5} \end{bmatrix}$$

which simplify as

$$\mathbf{x} \mapsto \begin{bmatrix} -\frac{1}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{1}{5} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \frac{2}{5} \\ \frac{4}{5} \end{bmatrix}; \qquad \mathbf{x} \mapsto \begin{bmatrix} \frac{4}{5} & -\frac{2}{5} \\ -\frac{2}{5} & -\frac{4}{5} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \frac{2}{5} \\ \frac{4}{5} \end{bmatrix}.$$

[11] Pumpkin image downloaded from https://openclipart.org/detail/86665/plain-pumpkin.



Solutions to Selected Exercises

Section 1.1

1a: The number of rows always comes first in the size of a matrix, so the matrix has 15 rows.

2b: The number of columns always comes second in the size of a matrix, so the matrix has 5 columns.

3c: A matrix with size 4×14 has 4 rows with 14 entries each (equivalently it has 14 columns with 4 entries each) for a total of $4 \cdot 14 = 56$ entries.

4d: $M_{3,1}$ means the entry of M in the third row and first column, so the answer is -2.

5d: $N_{:,3}$ means the third column of N, so its size is 5 rows by 1 column or 5×1.

5h: $N_{\backslash 4,2}$ means the submatrix formed by deleting row 4 and column 2 of N, which means $N_{\backslash 4,2}$ will have one fewer row and one fewer column than N making its size 4×5.

6b: $A_{6,:}$ means the sixth row of A so it is $\begin{bmatrix} 2 & -4 & 10 & -7 & -3 \end{bmatrix}$.

6f: $A_{2:4,2:3}$ means the submatrix of A containing the intersection of rows two through four with columns 2 through 3, so

$$A_{2:4,2:3} = \begin{bmatrix} -11 & 10 \\ -10 & 12 \\ -1 & 3 \end{bmatrix}.$$

Section 1.2

1i: Scalar products can always be computed, and is done so entry-wise. Each entry is multiplied by the scalar:

$$2\begin{bmatrix} -1 & 6 \\ 8 & 15 \end{bmatrix} = \begin{bmatrix} 2(-1) & 2(6) \\ 2(8) & 2(15) \end{bmatrix}$$

1n: Since these matrices are not the same size, there are entries in one that have no corresponding entry in the other. Therefore the difference is not defined. It cannot be computed.

6: In mathematics and logic a statement is either always true (true for all possible values of the variables) or it is false. Since there are matrices for which $M - N \neq N - M$ (see counterexample below) the statement is false.

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 4 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

9: SageMath uses calculator notation to do arithmetic computation, so $3A+4T$ is input as `3*A+4*T`. See SageMathCell 123. The result is

```
[ 286  188  -89 -110 -118 -132]
[-156   94   65  -30 -132  -58]
[ -51 -224 -195 -184 -324 -104]
[   6  281  120  112  122  -38]
[  22   -5  -51  -49  155  179]
```

Section 1.3

1d: For matrix multiplication to be defined, the left matrix must have the same number of columns (in this example it has one) as the right matrix has rows (in this example it has three). Therefore the matrix product $\begin{bmatrix} -9 \\ -4 \\ 4 \end{bmatrix}\begin{bmatrix} 2 \\ 9 \\ -6 \end{bmatrix}$ is undefined.

1d: Row-column multiplication is the sum of the entry-by-entry products (first entry times first entry plus second entry times second entry plus third entry times third entry):

$$\begin{aligned} \begin{bmatrix} -1 & 0 & -3 \end{bmatrix}\begin{bmatrix} 6 \\ -2 \\ 5 \end{bmatrix} &= -1(6) + 0(-2) + (-3)(5) \\ &= -6 + 0 - 15 = -21. \end{aligned}$$

2d: Because a column matrix has exactly one element per row and a row matrix one element per column, every column-matrix-row-matrix product is defined. The left matrix has the same number of columns (one) as the right matrix has rows (also one). The i,j-entry of the product is the product of entry in the i^{th} row of the left matrix with the entry in the j^{th} column of the right matrix. For example, the 1, 1-entry of the product is $(6.3)(2.3) = 14.49$ and the 2, 1-entry is $(4.1)(2.3) = 9.43$. Placing the right matrix just to the right and below the left matrix can help with the organization:

$$\begin{array}{cc} \begin{bmatrix} 6.3 \\ 4.1 \\ 3.4 \end{bmatrix} & \begin{bmatrix} 14.49 & 28.35 \\ 9.43 & 18.45 \\ 7.82 & 15.3 \end{bmatrix} \\ & \begin{bmatrix} 2.3 & 4.5 \end{bmatrix}. \end{array}$$

The answer is $\begin{bmatrix} 14.49 & 28.35 \\ 9.43 & 18.45 \\ 7.82 & 15.3 \end{bmatrix}$.

2e: In matrix multiplication, the left matrix must have the same number of columns (must be as wide) as the right matrix has rows (is tall). The left matrix of this example has 3 columns while the right matrix has 4 rows, so the product is undefined.

2f: The i, j-entry of a product is the product of the i^{th} row of the left matrix with the j^{th} column of the right matrix. Therefore the product will have as many rows as the left matrix and as many columns as the right matrix. In this example, that means 2 rows and 1 column.

$$1, 1\text{-entry: } \begin{bmatrix} -3 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} = 1$$

$$2, 1\text{-entry: } \begin{bmatrix} 2 & 5 & 7 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} = 45$$

Placing the right matrix just to the right and below the left matrix can help keep this straight:

$$\begin{array}{cc} \begin{bmatrix} -3 & 0 & 1 \\ 2 & 5 & 7 \end{bmatrix} & \begin{bmatrix} 1 \\ 45 \end{bmatrix} \\ & \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}. \end{array}$$

The answer is $\begin{bmatrix} 1 \\ 45 \end{bmatrix}$.

3e: $\mathbf{u}^T\mathbf{v} = -74$:

$$\mathbf{u}^T\mathbf{v} = \begin{bmatrix} 10 & 2 & -3 \end{bmatrix} \begin{bmatrix} -11 \\ 3 \\ -10 \end{bmatrix} = 10(-11) + 2(3) + (-3)(-10) = -110 + 6 + 30 = -74$$

3g: $\mathbf{v}^T\mathbf{u} = -74$:

$$\mathbf{v}^T\mathbf{u} = \begin{bmatrix} -11 & 3 & -10 \end{bmatrix} \begin{bmatrix} 10 \\ 2 \\ -3 \end{bmatrix} = (-11)10 + (3)2 + (-10)(-3) = -110 + 6 + 30 = -74$$

14a: Matrices or vectors may be used in SageMath.

`print((u.transpose()*v)[0,0])` produces

```
35150214
```

`print(vector(u)*vector(v))` produces

```
35150214
```

15: The transpose of a matrix is computed using the `.transpose()` method in SageMath.

$\left[Q^T R\right]$ `print(Q.transpose()*R)` produces

```
[ -571 -2517  -378  -941   100 -1176]
[-3588 -2891 -2430 -1283 -1422  -432]
[-2838 -1795 -2412 -1092 -1456   162]
[  -93  2587  1859  2531  1318   857]
[-2980 -2369  -957 -1053  -250  -379]
[ -660  1567  -678   708 -1417   514]
```

$\left[QR^T\right]$ `print(Q*R.transpose())` produces

```
[-2563  1613  1516 -1620  -280]
[ 1452   617 -5796  2035  1519]
[ 2529 -1187 -1066   650  1886]
[-1058  2668   575 -1211    85]
[  919  -140  -787  -221  1144]
```

They are not equal. Note they are not even the same size. $Q^T R$ is 6×6 while QR^T is 5×5.

Section 1.4

1f: $\|\mathbf{u}\|$ is the magnitude or norm of $\mathbf{u}$, which is defined as $\sqrt{\mathbf{u}^T\mathbf{u}}$. In this case,

$$\|\mathbf{u}\| = \sqrt{\begin{bmatrix} 2 & -6 & 12 \end{bmatrix}\begin{bmatrix} 2 \\ -6 \\ 12 \end{bmatrix}} = \sqrt{2^2 + (-6)^2 + 12^2} = \sqrt{4 + 36 + 144} = \sqrt{184}$$

2f: $d(\mathbf{u}, \mathbf{v})$ is the distance between $\mathbf{u}$ and $\mathbf{v}$, which is defined as the norm of their difference, $\|\mathbf{u} - \mathbf{v}\|$. In this case,

$$\|\mathbf{u} - \mathbf{v}\| = \left\| \begin{bmatrix} 2 \\ -6 \\ 12 \end{bmatrix} - \begin{bmatrix} -6 \\ -10 \\ 4 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 8 \\ 4 \\ 8 \end{bmatrix} \right\| = \sqrt{8^2 + 4^2 + 8^2} = \sqrt{144} = 12$$

3f: The simplest way to check whether two vectors are orthogonal is to check whether their dot product is zero. In this case,

$$\mathbf{u}^T\mathbf{v} = \begin{bmatrix} 2 & -6 & 12 \end{bmatrix}\begin{bmatrix} -6 \\ -10 \\ -4 \end{bmatrix} = 2(-6) + (-6)(-10) + 12(-4) = -12 + 60 - 48 = 0$$

Since the dot product is zero, the vectors are orthogonal.

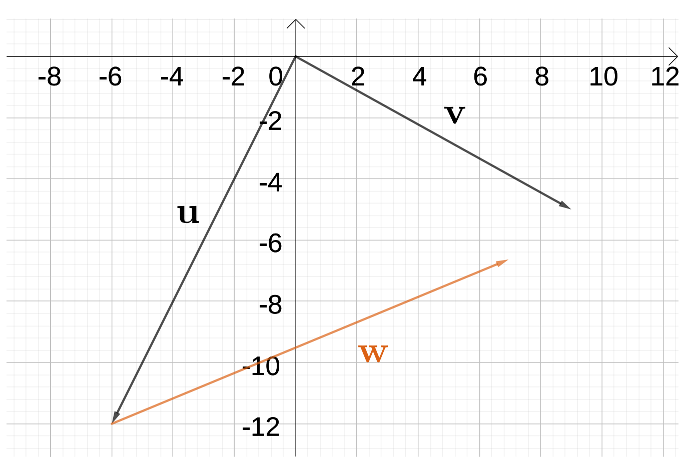
4c: In order for the vectors to be orthogonal their dot product must be zero. Setting the dot product equal to zero gives an equation that can be solved for k:

$$\begin{aligned} \begin{bmatrix} -2 & -6 & -3 \end{bmatrix}\begin{bmatrix} -7 \\ k \\ -10 \end{bmatrix} &= 0 \\ (-2)(-7) + (-6)k + (-3)(-10) &= 0 \\ 14 - 6k + 30 &= 0 \\ 44 &= 6k \\ \frac{22}{3} &= k \end{aligned}$$

so the solution is $k = \frac{22}{3}$.

5c: The sum of vectors is the vector with tail coinciding with the tail of the first addend and head coinciding with the head of the last addend. In this case, the tail of the first addend is at $(0, 0)$ and the head of the last addend is at $(5, 3)$, so the answer is the vector from $(0, 0)$ to $(5, 3)$: $\begin{bmatrix} 5 \\ 3 \end{bmatrix}$.

6c: Following the hint:



The orange vector **w** is the one requested. Since it is being added to **u**, its tail is placed at the head of **u**, and since the distance of this sum from **v** is supposed to be 1 unit from **v**, the head of **w** must land one unit from the head of **v**. There are many answers (none of which are shown in the diagram), for example placing the head of **w** at $(10, -5)$ we have

$$\mathbf{w} = \begin{bmatrix} 10 \\ -5 \end{bmatrix} - \begin{bmatrix} -6 \\ -12 \end{bmatrix} = \begin{bmatrix} 16 \\ 7 \end{bmatrix}.$$

Solutions may also be found algebraically. Plugging values of **u** and **v** into the given equation will allow solving for **w**:

$$\begin{aligned} d(\mathbf{u}+\mathbf{w},\mathbf{v}) &= 1 \\ \|\mathbf{u}+\mathbf{w}-\mathbf{v}\| &= 1 \\ \left\| \begin{bmatrix} -6 \\ -12 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} - \begin{bmatrix} 9 \\ -5 \end{bmatrix} \right\| &= 1 \\ \left\| \begin{bmatrix} w_1 - 15 \\ w_2 - 7 \end{bmatrix} \right\| &= 1 \\ \sqrt{(w_1-15)^2 + (w_2-7)^2} &= 1 \\ (w_1-15)^2 + (w_2-7)^2 &= 1 \\ (w_1-15)^2 &= 1 - (w_2-7)^2 \end{aligned} \tag{7.5.8}$$

There are infinitely many solutions. We may choose w_2 arbitrarily as long as $1 - (w_2-7)^2 \geq 0$ and solve for w_1. For example, $w_2 = 7.5$ and $w_1 = 15 + \sqrt{1 - .5^2} = 15 + \frac{\sqrt{3}}{2}$, giving $\begin{bmatrix} 15 + \frac{\sqrt{3}}{2} \\ \frac{15}{2} \end{bmatrix}$ as one solution. Note that the solution derived from the sketch satisfies (7.5.8) since $(16-15)^2 = 1 - (7-7)^2$.

10b: Yes. Their dot product is zero:

$$(-12.1\mathbf{u})^T(0.12\mathbf{v}) = -1.452\mathbf{u}^T\mathbf{v} = 0.$$

We do not need to have coordinates to draw this conclusion:

$$\begin{aligned} (-12.1\mathbf{u})^T(0.12\mathbf{v}) &= \left(-12.1 \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}\right)\left(0.12 \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}\right) \\ &= \left(\begin{bmatrix} -12.1u_1 & -12.1u_2 & \cdots & -12.1u_n \end{bmatrix}\right)\left(\begin{bmatrix} 0.12v_1 \\ 0.12v_2 \\ \vdots \\ 0.12v_n \end{bmatrix}\right) \\ &= (-12.1u_1)(0.12v_1) + (-12.1u_2)(0.12v_2) + \cdots + (-12.1u_n)(0.12v_n) \\ &= -1.452u_1v_1 - 1.452u_2v_2 - \cdots - 1.452u_nv_n \\ &= -1.452(u_1v_1 + u_2v_2 + \cdots + u_nv_n) \\ &= -1.452\mathbf{u}^T\mathbf{v} \end{aligned}$$

That they are orthogonal can also be argued geometrically. Scaling a vector does not change its direction, so scaling **u** and **v** does not change either of their directions. If they are orthogonal to begin with, they are orthogonal after scaling.

11: The code

```
v = D.row(2)
print(v.norm())
```

produces

```
sqrt(177)
```

Note the same result can be reached with the single line `print(D.row(2).norm())`.

Section 1.5

1c: According to formula (1.5.4),

$$C_{2,1} = (-1)^{2+1}\det(-6) = -\det(-6)$$

1e: According to formula (1.5.4),

$$C_{1,3} = (-1)^{1+3}\det\begin{pmatrix} -5 & -10 \\ -9 & 8 \end{pmatrix} = \det\begin{pmatrix} -5 & -10 \\ -9 & 8 \end{pmatrix}$$

2c: The determinant of a 1×1 matrix is the lone entry of the matrix, so $\det(30) = 30$.

2g: Using formula (1.5.1), or equivalently formula (1.5.3),

$$\begin{aligned} \begin{vmatrix} 18 & 5 \\ 14 & -16 \end{vmatrix} &= 18(-1)^{1+1}\det(-16) + 5(-1)^{1+2}\det(14) \\ &= 18(-16) - 5(14) = -358 \end{aligned}$$

2k: Using formula (1.5.1), or equivalently formula (1.5.3),

$$\begin{aligned} \det\begin{pmatrix} -3 & -1 & -9 \\ 1 & -4 & -8 \\ 2 & 9 & 6 \end{pmatrix} &= -3(-1)^{1+1}\det\begin{pmatrix} -4 & -8 \\ 9 & 6 \end{pmatrix} + (-1)(-1)^{1+2}\det\begin{pmatrix} 1 & -8 \\ 2 & 6 \end{pmatrix} \\ &\quad + (-9)(-1)^{1+3}\det\begin{pmatrix} 1 & -4 \\ 2 & 9 \end{pmatrix} \\ &= -3\,(-4(6) - (-8)(9)) + (1(6) - (-8)(2)) - 9\,(1(9) - (-4)(2)) \\ &= -3(48) + 22 - 9(17) = -275 \end{aligned}$$

2o: Using formula (1.5.1), or equivalently formula (1.5.3),

$$\begin{aligned} \det\begin{pmatrix} 5 & 0 & 2 & 8 \\ 4 & 8 & 6 & -2 \\ 0 & -1 & 6 & 0 \\ 0 & 3 & -1 & 3 \end{pmatrix} &= 5(-1)^{1+1}\det\begin{pmatrix} 8 & 6 & -2 \\ -1 & 6 & 0 \\ 3 & -1 & 3 \end{pmatrix} + 0 + 2(-1)^{1+3}\det\begin{pmatrix} 4 & 8 & -2 \\ 0 & -1 & 0 \\ 0 & 3 & 3 \end{pmatrix} \\ &\quad + 8(-1)^{1+4}\det\begin{pmatrix} 4 & 8 & 6 \\ 0 & -1 & 6 \\ 0 & 3 & -1 \end{pmatrix} \\ &= 5\left(8\det\begin{pmatrix} 6 & 0 \\ -1 & 3 \end{pmatrix} - 6\det\begin{pmatrix} -1 & 0 \\ 3 & 3 \end{pmatrix} - 2\det\begin{pmatrix} -1 & 6 \\ 3 & -1 \end{pmatrix}\right) \\ &\quad + 2\left(4\det\begin{pmatrix} -1 & 0 \\ 3 & 3 \end{pmatrix} - 8\det\begin{pmatrix} 0 & 0 \\ 0 & 3 \end{pmatrix} - 2\det\begin{pmatrix} 0 & -1 \\ 0 & 3 \end{pmatrix}\right) \\ &\quad - 8\left(4\det\begin{pmatrix} -1 & 6 \\ 3 & -1 \end{pmatrix} - 8\det\begin{pmatrix} 0 & 6 \\ 0 & -1 \end{pmatrix} + 6\det\begin{pmatrix} 0 & -1 \\ 0 & 3 \end{pmatrix}\right) \\ &= 5\,(8(18-0) - 6(-3-0) - 2(1-18)) \\ &\quad 2\,(4(-3-0) - 8(0-0) - 2(0-0)) \\ &\quad - 8\,(4(1-18) - 8(0-0) + 6(0-0)) \\ &= 5(196) + 2(-12) - 8(-68) = 1500 \end{aligned}$$

Section 1.6

1c: The 1×1 identity matrix is $\left[\begin{array}{c} 1 \end{array}\right]$, so we need a matrix M such that

$$M\left[\begin{array}{c} \frac{2}{\sqrt{3}} \end{array}\right] = \left[\begin{array}{c} \frac{2}{\sqrt{3}} \end{array}\right] M = \left[\begin{array}{c} 1 \end{array}\right].$$

There is no formula. We just have to recognize that the matrix in question is the 1×1 matrix with its lone entry equal to the reciprocal of $\frac{2}{\sqrt{3}}$:

$$\left[\begin{array}{c} \frac{2}{\sqrt{3}} \end{array}\right]^{-1} = \left[\begin{array}{c} \frac{\sqrt{3}}{2} \end{array}\right]$$

1g: Using formula (1.6.2),

$$\begin{aligned}
\left[\begin{array}{cc} 5 & -3 \\ -5 & 4 \end{array}\right]^{-1} &= \frac{1}{5(4)-(-3)(-5)}\left[\begin{array}{cc} C_{1,1} & C_{2,1} \\ C_{1,2} & C_{2,2} \end{array}\right] \\
&= \frac{1}{5}\left[\begin{array}{cc} 4 & 3 \\ 5 & 5 \end{array}\right] = \left[\begin{array}{cc} \frac{4}{5} & \frac{3}{5} \\ 1 & 1 \end{array}\right]
\end{aligned}$$

1i: $\left[\begin{array}{ccc} 2 & -3 & \sqrt{7} \\ 12 & \sqrt{28} & 5 \end{array}\right]$ is not a square matrix, so it does not have an inverse.

1m: Letting $M = \left[\begin{array}{ccc} 6 & 3 & 0 \\ -1 & -1 & 6 \\ 0 & 0 & 7 \end{array}\right]$ and using formula (1.6.2),

$$M^{-1} = \frac{1}{\det M}\left[\begin{array}{ccc} C_{1,1} & C_{2,1} & C_{3,1} \\ C_{1,2} & C_{2,2} & C_{3,2} \\ C_{1,3} & C_{2,3} & C_{3,3} \end{array}\right]$$

and

$$\begin{aligned}
\det M &= 6\det\begin{pmatrix} -1 & 6 \\ 0 & 7 \end{pmatrix} - 3\det\begin{pmatrix} -1 & 6 \\ 0 & 7 \end{pmatrix} = 6(-7) - 3(-7) = -21 \\
C_{1,1} &= \det\begin{pmatrix} -1 & 6 \\ 0 & 7 \end{pmatrix} = -7 \\
C_{2,1} &= -\det\begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix} = -21 \\
C_{3,1} &= \det\begin{pmatrix} 3 & 0 \\ -1 & 6 \end{pmatrix} = 18 \\
C_{1,2} &= -\det\begin{pmatrix} -1 & 6 \\ 0 & 7 \end{pmatrix} = 7
\end{aligned}$$

$$\begin{aligned}
C_{2,2} &= \det\begin{pmatrix} 6 & 0 \\ 0 & 7 \end{pmatrix} = 42 \\
C_{3,2} &= -\det\begin{pmatrix} 6 & 0 \\ -1 & 6 \end{pmatrix} = -36 \\
C_{1,3} &= \det\begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix} = 0 \\
C_{2,3} &= -\det\begin{pmatrix} 6 & 3 \\ 0 & 0 \end{pmatrix} = 0 \\
C_{3,3} &= \det\begin{pmatrix} 6 & 3 \\ -1 & -1 \end{pmatrix} = -6-(-3) = -3
\end{aligned}$$

Therefore

$$M^{-1} = \frac{1}{-21}\begin{bmatrix} -7 & -21 & 18 \\ 7 & 42 & -36 \\ 0 & 0 & -3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & 1 & -\frac{6}{7} \\ -\frac{1}{3} & -2 & \frac{12}{7} \\ 0 & 0 & \frac{1}{7} \end{bmatrix}$$

5: Applying equation (1.6.3) with $A = \begin{bmatrix} 1.4 & -70 \\ -29 & 95 \\ -12 & -43 \end{bmatrix}$ and $AB = \begin{bmatrix} 80 & 4.9 \\ -62 & -52 \\ -32 & 52 \end{bmatrix}$, it must be that

$$\begin{bmatrix} 80 & 4.9 \\ -62 & -52 \\ -32 & 52 \end{bmatrix} \cdot B^{-1} = (AB)B^{-1} = A = \begin{bmatrix} 1.4 & -70 \\ -29 & 95 \\ -12 & -43 \end{bmatrix}.$$

This operation can be thought of as right-multiplying both sides of

$$\begin{bmatrix} 1.4 & -70 \\ -29 & 95 \\ -12 & -43 \end{bmatrix} \cdot B = \begin{bmatrix} 80 & 4.9 \\ -62 & -52 \\ -32 & 52 \end{bmatrix}$$

by B^{-1}:

$$\begin{bmatrix} 1.4 & -70 \\ -29 & 95 \\ -12 & -43 \end{bmatrix} \cdot B \cdot B^{-1} = \begin{bmatrix} 80 & 4.9 \\ -62 & -52 \\ -32 & 52 \end{bmatrix} \cdot B^{-1}$$

which reduces to

$$\begin{bmatrix} 1.4 & -70 \\ -29 & 95 \\ -12 & -43 \end{bmatrix} \cdot I = \begin{bmatrix} 80 & 4.9 \\ -62 & -52 \\ -32 & 52 \end{bmatrix} \cdot B^{-1}$$

and finally to

$$\begin{bmatrix} 1.4 & -70 \\ -29 & 95 \\ -12 & -43 \end{bmatrix} = \begin{bmatrix} 80 & 4.9 \\ -62 & -52 \\ -32 & 52 \end{bmatrix} \cdot B^{-1}.$$

9: SageMathCell 124 One way to complete the code is

```
A = matrix(3,3,[7,-5,-2,-3,3,1,-3,2,1])
B = matrix(3,3,[1,2,2,2,8,7,-3,-5,-5])
# Compute (AB)^-1
print("(a)")
print((A*B).inverse()); print()
# Compute A^-1 B^-1
print("(b)")
print(A.inverse()*B.inverse()); print()
# Compute B^-1 A^-1
print("(c)")
print(B.inverse()*A.inverse())
```

which produces

```
(a)
[-11  -7 -17]
[-20 -13 -30]
[ 26  17  39]

(b)
[ -2   0  -1]
```

```
[-25   2  -7]
[ 58  -5  15]

(c)
[-11  -7 -17]
[-20 -13 -30]
[ 26  17  39]
```

10: SageMathCell 125 One way to code the computation is

```
encoder = matrix(3,3,[1,-4,-2,-3,7,3,0,2,1])
decoder = encoder.inverse()
message = matrix(3,5,[-589,-602,-244,-546,33,
                      861,958,224,768,-99,
                      339,317,180,325,0])
print("Encoded message:")
print(message); print()
print("Decoded message:")
print(decoder*message)
```

which produces

```
Encoded message:
[-589 -602 -244 -546   33]
[ 861  958  224  768  -99]
[ 339  317  180  325    0]

Decoded message:
[ 89  32 116 104  33]
[111 103  32 105   0]
[117 111 116 115   0]
```

and these numbers are ASCII codes for “You got this!”.

Section 1.7

1c: Since $\mathbf{v}$ is an eigenvector of A, it must be that $A\mathbf{v} = \lambda\mathbf{v}$ for some scalar λ. Computing $A\mathbf{v}$ will reveal the value of λ:

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

Since $\begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix} = 2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$, the eigenvalue λ must be 2.

2d: The characteristic polynomial of a square matrix M is $\det(M - \lambda I)$. In this case,

$$\begin{aligned} \det\left(\begin{bmatrix} -8 & -3 \\ 3 & -2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) &= \det \begin{pmatrix} -8-\lambda & -3 \\ 3 & -2-\lambda \end{pmatrix} \\ &= (-8-\lambda)(-2-\lambda) + 9. \end{aligned}$$

The characteristic polynomial is $(-8-\lambda)(-2-\lambda)+9$ and can be expanded to yield the standard form $\lambda^2+10\lambda+25$.

3d: The eigenvalues of a square matrix A are the roots of its characteristic polynomial. That is, solutions of the equation $\det(A - \lambda I) = 0$. In this case,

$$\begin{aligned}
\det\left(\begin{bmatrix} -7 & 25 \\ -1 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) &= 0 \\
\det\begin{pmatrix} -7-\lambda & 25 \\ -1 & 3-\lambda \end{pmatrix} &= 0 \\
(-7-\lambda)(3-\lambda) + 25 &= 0 \\
\lambda^2 + 4\lambda + 4 &= 0 \\
(\lambda+2)^2 &= 0 \\
\lambda &= -2
\end{aligned}$$

so this matrix has one eigenvalue, 2.

3k: The eigenvalues of a square matrix A are the roots of its characteristic polynomial. That is, solutions of the equation $\det(A - \lambda I) = 0$. In this case,

$$\begin{aligned}
\det\left(\begin{bmatrix} -3 & 0 & 0 \\ 4 & 13 & -12 \\ 4 & 16 & -15 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) &= 0 \\
\det\begin{pmatrix} -3-\lambda & 0 & 0 \\ 4 & 13-\lambda & -12 \\ 4 & 16 & -15-\lambda \end{pmatrix} &= 0 \\
(-3-\lambda)\det\begin{pmatrix} 13-\lambda & -12 \\ 16 & -15-\lambda \end{pmatrix} &= 0 \\
(-3-\lambda)\left((13-\lambda)(-15-\lambda) + 192\right) &= 0 \\
(-3-\lambda)(\lambda^2 + 2\lambda - 3) &= 0 \\
-\lambda^3 - 5\lambda^2 - 3\lambda + 9 &= 0.
\end{aligned}$$

Checking for integer roots first, the rational roots theorem says the only possibilities are factors of the constant term, 9 divided by factors of the leading coefficient, -1. That is, $\pm 1, \pm 3, \pm 9$. Starting with 1, $-(1)^3 - 5(1)^2 - 3(1) + 9 = -1 - 5 - 3 + 9 = 0$. How lucky, a hit on the first try! Factoring $\lambda - 1$ out of $-\lambda^3 - 5\lambda^2 - 3\lambda + 9$ using synthetic division:

$$\begin{array}{r|rrrr}
1 & -1 & -5 & -3 & 9 \\
 & & -1 & -6 & -9 \\ \hline
 & -1 & -6 & -9 & \boxed{0}
\end{array}$$

yields $-\lambda^3 - 5\lambda^2 - 3\lambda + 9 = (\lambda - 1)(-\lambda^2 - 6\lambda - 9)$. Completing the factoring (factoring the quadratic $-\lambda^2 - 6\lambda - 9$) gives the characteristic equation

$$\begin{aligned}
(\lambda - 1)(-\lambda - 3)(\lambda + 3) &= 0 \\
\lambda &= -3, 1
\end{aligned}$$

so this matrix has two eigenvalues, -3 and 1.

4c: The eigenpair $\lambda, \mathbf{v}$ satisfies the equation $A\mathbf{v} = \lambda\mathbf{v}$, which can be solved for $\mathbf{v}$. Letting $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$:

$$\begin{aligned}
\begin{bmatrix} -4 & 2 \\ -16 & 8 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} &= 0\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\
\begin{bmatrix} -4v_1 + 2v_2 \\ -16v_1 + 8v_2 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\end{aligned}$$

so v_1 and v_2 must satisfy the system

$$\begin{aligned} -4v_1 + 2v_2 &= 0 \\ -16v_1 + 8v_2 &= 0 \end{aligned}$$

Solving the first equation for v_2 as an attempt to solve the system by substitution: $2v_2 = 4v_1$ so $v_2 = 2v_1$. Substituting into the second equation, $-16v_1 + 8(2v_1) = 0$ yields $0 = 0$, a true statement for all values of v_1! This means v_1 can be anything and v_2 must be $2v_1$. For example,

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ is an eigenvector}$$

but any vector of the form $\begin{bmatrix} r \\ 2r \end{bmatrix}$ is a valid solution.

11d: SageMathCell 126 One possible solution is

```
M = matrix(2,2,[-8,-3,3,-2])
print(M); print()
print(M.charpoly())
```

which produces

```
[-8 -3]
[ 3 -2]

x^2 + 10*x + 25
```

The solution of 2d was $\lambda^2 + 10\lambda + 25$, which is the same except for the (dummy) variable, so is the same solution.

Section 2.1

3b: The linear system represented by the matrix is

$$\begin{aligned} 11v_1 &= 9 \\ 5v_2 &= -7 \\ v_3 &= -13 \\ -2v_4 &= 6 \end{aligned}$$

so the solution is $v_1 = \frac{9}{11}$, $v_2 = -\frac{7}{5}$, $v_3 = -13$, $v_4 = -3$.

3d: The linear system represented by the matrix is

$$\begin{aligned} 11v_1 + 9v_3 &= 12 \\ -8v_2 - 4v_3 &= -1 \\ v_3 &= 2 \end{aligned}$$

Starting with $v_3 = 2$ and substituting into the other equations yields $-8v_2 - 4(2) = -1$ so $v_2 = -\frac{7}{8}$ and $11v_1 + 9(2) = 12$ so $v_1 = -\frac{6}{11}$. The solution is therefore $v_1 = -\frac{6}{11}$, $v_2 = -\frac{7}{8}$, $v_3 = 2$.

9c: Adding 4 times row 2 to row 3 of the identity matrix yields the given matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{4I_{2,:}+I_{3,:}\to I_{3,:}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$$

so the elementary row operation must be adding 4 times row 2 to row 3.

12b: The first row of the left matrix holds the coefficients of the linear combination so the first row of the product must be

$$-2\begin{bmatrix} 4 & -3 & 2 \end{bmatrix} + 1\begin{bmatrix} 12 & 8 & -5 \end{bmatrix} = \begin{bmatrix} -8+12 & 6+8 & -4-5 \end{bmatrix} = \begin{bmatrix} 4 & 14 & -9 \end{bmatrix}$$

Section 2.2

1a: Answers will vary depending on the steps taken. For a matrix with two rows and no columns of zeros, the only requirement of row echelon form is a 0 in the 2,1-entry. One way to reduce is by first scaling the two rows and then replacing the second row:

$$\begin{bmatrix} -2 & -4 & -10 \\ -5 & 2 & -1 \end{bmatrix} \xrightarrow[2A_{2,:}\to A_{2,:}]{-5A_{1,:}\to A_{1,:}} \begin{bmatrix} 10 & 20 & 50 \\ -10 & 4 & -2 \end{bmatrix} \xrightarrow{A_{1,:}+A_{2,:}\to A_{2,:}} \begin{bmatrix} 10 & 20 & 50 \\ 0 & 24 & 48 \end{bmatrix}$$

1d: Answers will vary depending on the steps taken. For a matrix with three rows and no columns of zeros, row echelon form requires zeros in the 2,1-, 3,1-, and 3,2-entries. One way to reduce is as follows:

$$\begin{bmatrix} -1 & 0 & 2 \\ 4 & 1 & -3 \\ 1 & 1 & 3 \end{bmatrix} \xrightarrow[A_{1,:}+A_{3,:}\to A_{3,:}]{4A_{1,:}+A_{2,:}\to A_{2,:}} \begin{bmatrix} -1 & 0 & 2 \\ 0 & 1 & 5 \\ 0 & 1 & 5 \end{bmatrix} \xrightarrow{-A_{2,:}+A_{3,:}\to A_{3,:}} \begin{bmatrix} -1 & 0 & 2 \\ 0 & 1 & 5 \\ 0 & 0 & 0 \end{bmatrix}$$

2a: Reduced row echelon form requires ones in the pivot positions and zeros above. Beginning with the echelon form of question 1a, the pivot positions are the 1,1- and 2,2-entries.

$$\begin{bmatrix} 10 & 20 & 50 \\ 0 & 24 & 48 \end{bmatrix} \xrightarrow[\frac{1}{24}A_{2,:}\to A_{2,:}]{\frac{1}{10}A_{1,:}\to A_{1,:}} \begin{bmatrix} 1 & 2 & 5 \\ 0 & 1 & 2 \end{bmatrix} \xrightarrow{-2A_{2,:}+A_{1,:}\to A_{1,:}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

2d: Reduced row echelon form requires ones in the pivot positions and zeros above. Beginning with the echelon form of question 1a, the pivot positions are the 1,1- and 2,2-entries. The only thing remaining to do is get a one in the 1,1-entry.

$$\begin{bmatrix} -1 & 0 & 2 \\ 0 & 1 & 5 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{-A_{1,:}\to A_{1,:}} \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 5 \\ 0 & 0 & 0 \end{bmatrix}$$

3a: A homogeneous system has zero constants, so the associated system is

$$\begin{array}{rcrcr} 3x_1 & - & x_2 & = & 0 \\ & & 5x_2 & = & 0 \end{array}$$

The second equation requires $x_2 = 0$ and substituting this value into the first equation reveals that $x_1 = 0$ also. There are no nontrivial solutions.

3f: A homogeneous system has zero constants, so the associated system is

$$\begin{array}{rcrcrcr} 5x_1 & & & - & 3x_3 & = & 0 \\ & & -4x_2 & + & x_3 & = & 0 \\ & & & & 0 & = & 0 \end{array}$$

Answers will vary since there are infinitely many solutions, $x_1 = 12$, $x_2 = 5$, $x_3 = 20$ is one example. The third equation is always true (no matter the values of x_1, x_2, x_3). The second equation requires $x_2 = \frac{1}{4}x_3$ and the first equation requires $x_1 = \frac{3}{5}x_3$. Any solution where these requirements of x_1 and x_2 are met will suffice. For example, choosing $x_3 = 20$, we get $x_2 = 5$ and $x_1 = 12$.

4b: Putting the coefficients in an augmented matrix and row reducing to reduced row echelon form:

$$\begin{bmatrix} 1 & 4 & -4 \\ -3 & -11 & -5 \end{bmatrix} \xrightarrow{3A_{1,:}+A_{2,:}\to A_{2,:}} \begin{bmatrix} 1 & 4 & -4 \\ 0 & 1 & -17 \end{bmatrix} \xrightarrow{-4A_{2,:}+A_{1,:}\to A_{1,:}} \begin{bmatrix} 1 & 0 & 64 \\ 0 & 1 & -17 \end{bmatrix}$$

The reduced row echelon form represents the system $v_1 = 64$, $v_2 = -17$ (the solution).

4h: Putting the coefficients in an augmented matrix and row reducing to reduced row echelon form:

$$\begin{bmatrix} -3 & -35 & 10 & 2 \\ 9 & 130 & -40 & 2 \\ 9 & 120 & -35 & -4 \end{bmatrix} \xrightarrow[3A_{1,:}+A_{3,:}\to A_{3,:}]{3A_{1,:}+A_{2,:}\to A_{2,:}} \begin{bmatrix} -3 & -35 & 10 & 2 \\ 0 & 25 & -10 & 8 \\ 0 & 15 & -5 & 2 \end{bmatrix} \xrightarrow[-5A_{3,:}\to A_{3,:}]{3A_{2,:}\to A_{2,:}} \begin{bmatrix} -3 & -35 & 10 & 2 \\ 0 & 75 & -30 & 24 \\ 0 & -75 & 25 & -10 \end{bmatrix}$$

$$\xrightarrow{A_{2,:}+A_{3,:}\to A_{3,:}} \begin{bmatrix} -3 & -35 & 10 & 2 \\ 0 & 75 & -30 & 24 \\ 0 & 0 & -5 & 14 \end{bmatrix} \xrightarrow[2A_{3,:}+A_{1,:}\to A_{1,:}]{-6A_{3,:}+A_{2,:}\to A_{2,:}} \begin{bmatrix} -3 & -35 & 0 & 30 \\ 0 & 75 & 0 & -60 \\ 0 & 0 & -5 & 14 \end{bmatrix} \xrightarrow{\frac{1}{15}A_{2,:}\to A_{2,:}}$$

$$\begin{bmatrix} -3 & -35 & 0 & 30 \\ 0 & 5 & 0 & -4 \\ 0 & 0 & -5 & 14 \end{bmatrix} \xrightarrow{7A_{2,:}+A_{1,:}\to A_{1,:}} \begin{bmatrix} -3 & 0 & 0 & 2 \\ 0 & 5 & 0 & -4 \\ 0 & 0 & -5 & 14 \end{bmatrix}$$

Scaling each row appropriately produces

$$\begin{bmatrix} 1 & 0 & 0 & -\frac{2}{3} \\ 0 & 1 & 0 & -\frac{4}{5} \\ 0 & 0 & 1 & -\frac{14}{5} \end{bmatrix}$$

from which the solution is clearly $v_1 = -\frac{2}{3}$, $v_2 = -\frac{4}{5}$, $v_3 = -\frac{14}{5}$.

Section 2.3

4b: Begin by reducing:

$$\begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & -7 \\ 0 & -2 & 7 \end{bmatrix} \xrightarrow[A_{2,:}+A_{3,:}\to A_{3,:}]{-2A_{2,:}+A_{1,:}\to A_{1,:}} \begin{bmatrix} 2 & 0 & 19 \\ 0 & 2 & -7 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[\frac{1}{2}A_{2,:}\to A_{2,:}]{\frac{1}{2}A_{1,:}\to A_{1,:}} \begin{bmatrix} 1 & 0 & \frac{19}{2} \\ 0 & 1 & -\frac{7}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

which means x_3 is a free variable and the solution is

$$x_1 = -\frac{19}{2}x_3; \quad x_2 = \frac{7}{2}x_3.$$

In parametric vector form, using r for the arbitrary parameter:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = r \begin{bmatrix} -\frac{19}{2} \\ \frac{7}{2} \\ 1 \end{bmatrix}.$$

Equivalently,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = s \begin{bmatrix} -19 \\ 7 \\ 2 \end{bmatrix}.$$

5c: Eigenvectors are solutions of the system $(A - \lambda I)\mathbf{v} = 0$, for each lambda.

$\lambda = -6$**:** $A - (-6)I = \begin{bmatrix} 12 & -4 & 16 \\ 3 & -1 & 4 \\ -6 & 2 & -8 \end{bmatrix}$ which reduces as follows:

$$\begin{bmatrix} 12 & -4 & 16 \\ 3 & -1 & 4 \\ -6 & 2 & -8 \end{bmatrix} \xrightarrow{\frac{1}{4}A_{1,:}\to A_{1,:}} \begin{bmatrix} 3 & -1 & 4 \\ 3 & -1 & 4 \\ -6 & 2 & -8 \end{bmatrix} \xrightarrow[2A_{1,:}+A_{3,:}\to A_{3,:}]{-A_{1,:}+A_{2,:}\to A_{2,:}} \begin{bmatrix} 3 & -1 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}^T$ must satisfy only $3v_1 - v_2 + 4v_3 = 0$. v_2 and v_3 are free and $v_1 = \frac{1}{3}v_2 - \frac{4}{3}v_3$. In parametric vector form,

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = r \begin{bmatrix} 1/3 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} -4/3 \\ 0 \\ 1 \end{bmatrix}.$$

$\lambda = -3$: $A - (-3)I = \begin{bmatrix} 9 & -4 & 16 \\ 3 & -4 & 4 \\ -6 & 2 & -11 \end{bmatrix}$ which reduces as follows:

$$\begin{bmatrix} 9 & -4 & 16 \\ 3 & -4 & 4 \\ -6 & 2 & -11 \end{bmatrix} \xrightarrow{A_{1,:} \leftrightarrow A_{2,:}} \begin{bmatrix} 3 & -4 & 4 \\ 9 & -4 & 16 \\ -6 & 2 & -11 \end{bmatrix} \xrightarrow[2A_{1,:}+A_{3,:}\to A_{3,:}]{-3A_{1,:}+A_{2,:}\to A_{2,:}} \begin{bmatrix} 3 & -4 & 4 \\ 0 & 8 & 4 \\ 0 & -6 & -3 \end{bmatrix} \xrightarrow[-\frac{1}{3}A_{3,:}\to A_{3,:}]{\frac{1}{4}A_{2,:}\to A_{2,:}}$$
$$\begin{bmatrix} 3 & -4 & 4 \\ 0 & 2 & 1 \\ 0 & 2 & 1 \end{bmatrix} \xrightarrow{-A_{2,:}+A_{3,:}\to A_{3,:}} \begin{bmatrix} 3 & -4 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{2A_{2,:}+A_{1,:}\to A_{1,:}} \begin{bmatrix} 3 & 0 & 6 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[\frac{1}{2}A_{2,:}\to A_{2,:}]{\frac{1}{3}A_{1,:}\to A_{1,:}} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

Hence $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}^T$ must satisfy $v_1 = -2v_3$ and $v_2 = -\frac{1}{2}v_3$. v_3 is free. In parametric vector form,

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = t \begin{bmatrix} -2 \\ -\frac{1}{2} \\ 1 \end{bmatrix}.$$

In summary, eigenvectors take the form $r \begin{bmatrix} 1/3 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} -4/3 \\ 0 \\ 1 \end{bmatrix}$ or the form $t \begin{bmatrix} -2 \\ -\frac{1}{2} \\ 1 \end{bmatrix}$.

14: One way to complete the code is as follows.

```
M=matrix(5,5,[2049,-4548,-511,-5177,6023,-4526,10252,916,11438,
              -13292,-6947,15538,1740,17601,-20614,-1388,2866,
              263,3166,-3697,-5781,12812,1211,14321,-16671])
print("M ="); print(M); print()
# Find a row echelon form (but not reduced row echelon form)
print("Row echelon form:")
print(M.echelon_form()); print()
# Find the reduced row echelon form
print("Reduced row echelon form:")
print(M.rref())
```

which produces:

```
M =
[  2049  -4548   -511  -5177   6023]
[ -4526  10252    916  11438 -13292]
[ -6947  15538   1740  17601 -20614]
[ -1388   2866    263   3166  -3697]
[ -5781  12812   1211  14321 -16671]

Row echelon form:
[   1   38  102  149 -184]
[   0  134   67  402 -201]
[   0    0  134  268 -134]
[   0    0    0  804  268]
[   0    0    0    0    0]

Reduced row echelon form:
[   1    0    0    0 -1/3]
[   0    1    0    0 -5/3]
[   0    0    1    0 -5/3]
[   0    0    0    1  1/3]
[   0    0    0    0    0]
```

Section 3.1

2b: For any $r \times s$ matrix L and any $s \times t$ matrix R, the i,j-entry of LR is

$$(LR)_{i,j} = L_{i,1}R_{1,j} + L_{i,2}R_{2,j} + \cdots + L_{i,s}R_{s,j}. \tag{7.5.9}$$

Now suppose A is $\ell \times m$, B is $m \times n$, and C is $n \times p$. Then applying formula (7.5.9) with various substitutions for matrices L and R and counts s:

$$\begin{aligned}(A(BC))_{i,j} &= A_{i,1}(BC)_{1,j} + A_{i,2}(BC)_{2,j} + \cdots + A_{i,m}(BC)_{m,j} \\ &= A_{i,1}\left(B_{1,1}C_{1,j} + B_{1,2}C_{2,j} + \cdots + B_{1,n}C_{n,j}\right) \\ &\quad + A_{i,2}\left(B_{2,1}C_{1,j} + B_{2,2}C_{2,j} + \cdots + B_{2,n}C_{n,j}\right) \\ &\quad + \cdots + A_{i,m}\left(B_{m,1}C_{1,j} + B_{m,2}C_{2,j} + \cdots + B_{m,n}C_{n,j}\right)\end{aligned} \tag{7.5.10}$$

and

$$\begin{aligned}((AB)C)_{i,j} &= (AB)_{i,1}C_{1,j} + (AB)_{i,2}C_{2,j} + \cdots + (AB)_{i,n}C_{n,j} \\ &= (A_{i,1}B_{1,1} + A_{i,2}B_{2,1} + \cdots + A_{i,m}B_{m,1})\, C_{1,j} \\ &\quad + (A_{i,1}B_{1,2} + A_{i,2}B_{2,2} + \cdots + A_{i,m}B_{m,2})\, C_{2,j} \\ &\quad + \cdots + (A_{i,1}B_{1,n} + A_{i,2}B_{2,n} + \cdots + A_{i,m}B_{m,n})\, C_{n,j}\end{aligned} \tag{7.5.11}$$

By inspection, (7.5.10) and (7.5.11) both contain the mn terms

$$A_{i,x}B_{x,y}C_{y,j} \quad x = 1, 2, \ldots, m \text{ and } y = 1, 2, \ldots, n$$

and therefore are equal.

2g: For the arbitrary matrix M, $\left(M^T\right)_{i,j} = M_{j,i}$, which follows from the definition of the transpose. Applying this observation twice,

$$\left(\left(A^T\right)^T\right)_{i,j} = \left(A^T\right)_{j,i} = A_{i,j}.$$

3a: By the definition of inverse, (1.6.1), it must be shown that the product of the two matrices in either order is the identity.

To show that $(B^{-1}A^{-1})(AB) = I$:

$$\begin{aligned}(B^{-1}A^{-1})(AB) &= \left((B^{-1}A^{-1})A\right)B && \text{theorem 2 claim 4} \\ &= \left(B^{-1}(A^{-1}A)\right)B && \text{theorem 2 claim 4} \\ &= \left(B^{-1}I\right)B && \text{definition of inverse, (1.6.1)} \\ &= B^{-1}B && \text{theorem 2 claim 4} \\ &= I && \text{definition of inverse, (1.6.1)}\end{aligned}$$

To show that $(AB)(B^{-1}A^{-1}) = I$ is similar:

$$\begin{aligned}(AB)(B^{-1}A^{-1}) &= A\left(B(B^{-1}A^{-1})\right) && \text{theorem 2 claim 4} \\ &= A\left((BB^{-1})A^{-1}\right) && \text{theorem 2 claim 4} \\ &= A\left(IA^{-1}\right) && \text{definition of inverse, (1.6.1)} \\ &= AA^{-1} && \text{theorem 2 claim 4} \\ &= I && \text{definition of inverse, (1.6.1)}\end{aligned}$$

6: According to theorem 3 part 1, $3A + 7A = 10A$, so the calculation can be done without calculating $3A$ or $7A$ like so:

$$3A + 7A = 10A = 10\begin{bmatrix} 11 & 10 & -6 \\ -3 & 6 & 7 \end{bmatrix} = \begin{bmatrix} 110 & 100 & -60 \\ -30 & 60 & 70 \end{bmatrix}$$

You can check that $3A + 7A = \begin{bmatrix} 110 & 100 & -60 \\ -30 & 60 & 70 \end{bmatrix}$ by calculating $3A$ and $7A$ and adding them.

Section 3.2

1c: One way to proceed is to subtract from both sides and then multiply both sides by -1:

$$\begin{aligned}\begin{bmatrix} 0 & -19 \\ -1 & 8 \end{bmatrix} - \begin{bmatrix} 0 & -19 \\ -1 & 8 \end{bmatrix} - X &= \begin{bmatrix} -2 & 19 \\ -14 & 20 \end{bmatrix} - \begin{bmatrix} 0 & -19 \\ -1 & 8 \end{bmatrix} \\ -X &= \begin{bmatrix} -2 & 38 \\ -13 & 12 \end{bmatrix} \\ -1\,(-X) &= -1\left(\begin{bmatrix} -2 & 38 \\ -13 & 12 \end{bmatrix}\right) \\ X &= \begin{bmatrix} 2 & -38 \\ 13 & -12 \end{bmatrix}\end{aligned}$$

2b: **Left**-multiply both sides by a multiplicative inverse. There is no division of matrices.

$$\begin{aligned}\begin{bmatrix} 5 & 2 \\ 6 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 5 & 2 \\ 6 & 3 \end{bmatrix} X &= \begin{bmatrix} 5 & 2 \\ 6 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 13 & -13 \\ -19 & 7 \end{bmatrix} \\ IX &= \frac{1}{3}\begin{bmatrix} 3 & -2 \\ -6 & 5 \end{bmatrix}\begin{bmatrix} 13 & -13 \\ -19 & 7 \end{bmatrix} \\ X &= \frac{1}{3}\begin{bmatrix} 77 & -53 \\ -173 & 113 \end{bmatrix} = \begin{bmatrix} \frac{77}{3} & -\frac{53}{3} \\ -\frac{173}{3} & \frac{113}{3} \end{bmatrix}\end{aligned}$$

3d: **Right**-multiply both sides by P and **left**-multiply both sides by P^{-1}:

$$\begin{aligned}\left(PDP^{-1}\right)P &= AP \\ (PD)(P^{-1}P) &= AP \\ PD &= AP \\ P^{-1}(PD) &= P^{-1}AP \\ (P^{-1}P)D &= P^{-1}AP \\ D &= P^{-1}AP\end{aligned}$$

Since P^{-1} appears in the equation being solved, it is assumed to exist.

6c: The second row of a matrix product is the linear combination of the rows of the righthand matrix with coefficients coming from the second row of the lefthand matrix. In symbols, $(AB)_{2,:} = A_{2,1}B_{1,:} + A_{2,2}B_{2,:} + \cdots + A_{2,n}B_{n,:}$ (assuming A has n columns and B has n rows). Applied to this question,

$$\begin{aligned}&0\begin{bmatrix} -3 & 3 \end{bmatrix} - 3\begin{bmatrix} 2 & -4 \end{bmatrix} + 3\begin{bmatrix} -5 & 5 \end{bmatrix} + 2\begin{bmatrix} 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \end{bmatrix} + \begin{bmatrix} -6 & 12 \end{bmatrix} + \begin{bmatrix} -15 & 15 \end{bmatrix} + \begin{bmatrix} 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} -21 & 29 \end{bmatrix}\end{aligned}$$

7c: The third row of a matrix product is the linear combination of the rows of the righthand matrix with coefficients coming from the third row of the lefthand matrix. In symbols, $(AB)_{3,:} = A_{3,1}B_{1,:} + A_{3,2}B_{2,:} + \cdots + A_{3,n}B_{n,:}$ (assuming A has n columns and B has n rows). Applied to this question,

$$\begin{aligned}&4\begin{bmatrix} -3 & 3 \end{bmatrix} + 1\begin{bmatrix} 2 & -4 \end{bmatrix} + 4\begin{bmatrix} -5 & 5 \end{bmatrix} - 5\begin{bmatrix} 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -12 & 12 \end{bmatrix} + \begin{bmatrix} 2 & -4 \end{bmatrix} + \begin{bmatrix} -20 & 20 \end{bmatrix} + \begin{bmatrix} 0 & -5 \end{bmatrix} \\ &= \begin{bmatrix} -30 & 23 \end{bmatrix}\end{aligned}$$

8c: The second column of a matrix product is the linear combination of the columns of the lefthand matrix with coefficients coming from the second column of the righthand matrix. In symbols, $(AB)_{:,2} = A_{:,1}B_{1,2} + A_{:,2}B_{2,2} + \cdots + A_{:,n}B_{n,2}$ (assuming A has n columns and B has n rows). Applied to this question,

$$3\begin{bmatrix} -4 \\ 0 \\ 4 \end{bmatrix} - 4\begin{bmatrix} 5 \\ -3 \\ 1 \end{bmatrix} + 5\begin{bmatrix} -3 \\ 3 \\ 4 \end{bmatrix} + 1\begin{bmatrix} -4 \\ 2 \\ -5 \end{bmatrix} = \begin{bmatrix} -12 \\ 0 \\ 12 \end{bmatrix} + \begin{bmatrix} -20 \\ 12 \\ -4 \end{bmatrix} + \begin{bmatrix} -15 \\ 15 \\ 20 \end{bmatrix} + \begin{bmatrix} -4 \\ 2 \\ -5 \end{bmatrix} = \begin{bmatrix} -51 \\ 29 \\ 23 \end{bmatrix}$$

9c: The product has no third column since the righthand matrix has no third column.

Section 3.3

1a: (solution 1): If the vectors are augmented to form a matrix, then theorem 5 applies. It gives 6 ways to show that the columns of a matrix are linearly independent, parts (ii)-(vii). If we can show any one of them true, we have shown that the columns of the augmented matrix, which are the given vectors, are linearly independent. The simplest route to a conclusion is to use part (iv) of the theorem—M has a pivot position in every column—as determining pivot positions amounts to doing some row reduction.

Augmenting the vectors gives the matrix

$$M = \begin{bmatrix} -1 & 5 \\ -1 & 4 \end{bmatrix}.$$

Note that the vectors have been augmented in an order that makes row reduction simple (a -1 in the 1,1-entry). This is consistent with the idea that linear independence is a characterisitic of a set, where order of elements does not matter. The row reduction can be completed in one operation: add -1 times the first row to the second row, which yields

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

At this point, it is clear the matrix has two pivot positions, one in each column. By theorem 5 the columns of M are linearly independent.

Remark: We can also see at this point that M has no free variables—part (vii) of theorem 5—giving another way to conclude that the columns of M are linearly independent.

(solution 2): The definition of linear independence can be used just as well. The definition revolves around linear combinations of the vectors that sum to zero:

$$x_1\begin{bmatrix} 5 \\ 4 \end{bmatrix} + x_2\begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which can also be written as

$$\begin{bmatrix} 5 & -1 \\ 4 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

or

$$\begin{bmatrix} -1 & 5 \\ -1 & 4 \end{bmatrix}\begin{bmatrix} x_2 \\ x_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{7.5.12}$$

As in solution 1, we choose to set up the system to make the row reduction simple. Writing the augmented matrix for (7.5.12) and reducing:

$$\begin{bmatrix} -1 & 5 & 0 \\ -1 & 4 & 0 \end{bmatrix} \to \begin{bmatrix} -1 & 5 & 0 \\ 0 & -1 & 0 \end{bmatrix} \to \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

which means the (one and only) solution of the system is $x_1 = x_2 = 0$. By definition, the vectors are linearly independent.

1c: (solution 1): If the vectors are augmented to form a matrix, then theorem 5 applies. It gives 6 ways to show that the columns of a matrix are linearly independent, parts (ii)-(vii). If we can show any one of them true, we have shown that the columns of the augmented matrix, which are the given vectors, are linearly independent. The simplest route to a conclusion is to use part (iv) of the theorem—M has a pivot position in every column—as determining pivot positions amounts to doing some row reduction.

Augmenting the vectors gives the matrix

$$M = \begin{bmatrix} -1 & 4 \\ 2 & -3 \\ -1 & -3 \end{bmatrix}.$$

Note that the vectors have been augmented in an order that makes row reduction simple (a -1 in the 1,1-entry). This is consistent with the idea that linear independence is a characterisitic of a set, where order of elements does not matter. Enough row reduction can be completed in just two row operations:

$$\begin{bmatrix} -1 & 4 \\ 2 & -3 \\ -1 & -3 \end{bmatrix} \to \begin{bmatrix} -1 & 4 \\ 0 & 5 \\ 0 & -7 \end{bmatrix}.$$

At this point, it is clear the matrix has two pivot positions, one in each column. By theorem 5 the columns of M are linearly independent.

Remark: We can also see at this point that M has no free variables—part (vii) of theorem 5—giving another way to conclude that the columns of M are linearly independent.

(solution 2): The definition of linear independence can be used just as well. The definition revolves around linear combinations of the vectors that sum to zero:

$$x_1 \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ -3 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

which can also be written as

$$\begin{bmatrix} -1 & 4 \\ 2 & -3 \\ -1 & -3 \end{bmatrix} \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{7.5.13}$$

As in solution 1, we choose to set up the system to make the row reduction simple. Writing the augmented matrix for (7.5.13) and reducing:

$$\begin{bmatrix} -1 & 4 & 0 \\ 2 & -3 & 0 \\ -1 & -3 & 0 \end{bmatrix} \to \begin{bmatrix} -1 & 4 & 0 \\ 0 & 5 & 0 \\ 0 & -7 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & -4 & 0 \\ 0 & 1 & 0 \\ 0 & -7 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which means the (one and only) solution of the system is $x_1 = x_2 = 0$. By definition, the vectors are linearly independent.

2c: We are trying to conclude that the system has at most one solution for any constants. Part (vi) of theorem 5 is exactly this statement, which means if we can show that any one of the other conditions of theorem 5 holds, we are done. Parts (iii), (iv), and (vii) are within reach. Each one follows from row reduction of the coefficient matrix of the system. To make the work a little easier, we rewrite the system as

$$\begin{array}{rcrcrcl} -v_3 & - & v_2 & - & 2v_1 & = & b_1 \\ & + & 2v_2 & + & 7v_1 & = & b_2 \\ v_3 & - & v_2 & - & 4v_1 & = & b_3 \end{array}$$

and reduce the corresponding coefficient matrix:

$$\begin{bmatrix} -1 & -1 & -2 \\ 0 & 2 & 7 \\ 1 & -1 & -4 \end{bmatrix} \to \begin{bmatrix} -1 & -1 & -2 \\ 0 & 2 & 7 \\ 0 & -2 & -6 \end{bmatrix} \to \begin{bmatrix} -1 & -1 & -2 \\ 0 & 2 & 7 \\ 0 & 0 & 1 \end{bmatrix}$$

At this point it is clear that the system has no free variables (and that the coefficient matrix has a pivot in every column), so theorem 5 gives us that the system has at most one solution for any selection of constants.

2d: We are trying to conclude that the system has at most one solution for any constants. Part (vi) of theorem 5 is exactly this statement, which means if we can show that any one of the other conditions of theorem 5 holds, we are done. Parts (iii), (iv), and (vii) are within reach. Each one follows from row reduction of the coefficient matrix of the system. To make the work a little easier, we rewrite the system as

$$\begin{array}{rcrcrcl} -z & + & 3y & - & 6x & = & b_3 \\ 2z & + & 7y & + & 6x & = & b_2 \\ 7z & + & 5y & + & x & = & b_1 \\ & + & y & + & 5x & = & b_4 \end{array}$$

and reduce the corresponding coefficient matrix:

$$\begin{bmatrix} -1 & 3 & -6 \\ 2 & 7 & 6 \\ 7 & 5 & 1 \\ 0 & 1 & 5 \end{bmatrix} \to \begin{bmatrix} -1 & 3 & -6 \\ 0 & 13 & -6 \\ 0 & 26 & -41 \\ 0 & 1 & 5 \end{bmatrix} \to \begin{bmatrix} -1 & 3 & -6 \\ 0 & 1 & 5 \\ 0 & 26 & -41 \\ 0 & 13 & -6 \end{bmatrix} \to \begin{bmatrix} -1 & 3 & -6 \\ 0 & 1 & 5 \\ 0 & 0 & -171 \\ 0 & 0 & -71 \end{bmatrix}$$

At this point it is clear that the system has no free variables (and that the coefficient matrix has a pivot in every column), so theorem 5 gives us that the system has at most one solution for any selection of constants.

3a: We are asked to show that the homogeneous system has only the trivial solution. Part (iii) of theorem 5 is exactly this statement, which means if we can show that any one of the other conditions of theorem 5 holds, we are done. Parts (iv) and (vii) are within reach. Each one follows from row reduction of the coefficient matrix of the system:

$$\begin{bmatrix} 1 & 8 \\ 1 & -5 \end{bmatrix} \to \begin{bmatrix} 1 & 8 \\ 0 & -13 \end{bmatrix}$$

At this point it is clear that the system has no free variables (and that the coefficient matrix has a pivot in every column), so theorem 5 gives us that the homogeneous system has only the trivial solution.

3g: We are asked to show that the homogeneous system has only the trivial solution. Part (iii) of theorem 5 is exactly this statement, which means if we can show that any one of the other conditions of theorem 5 holds, we are done. Parts (iv) and (vii) are within reach. Each one follows from row reduction of the coefficient matrix of the system:

$$\begin{bmatrix} 6 & 3 & -1 \\ 5 & 0 & 1 \\ 1 & -4 & 1 \\ 5 & 7 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & -4 & 1 \\ 5 & 0 & 1 \\ 6 & 3 & -1 \\ 5 & 7 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & -4 & 1 \\ 0 & 20 & -4 \\ 0 & 27 & -7 \\ 0 & 27 & -9 \end{bmatrix} \to \begin{bmatrix} 1 & -4 & 1 \\ 0 & 20 & -4 \\ 0 & 27 & -7 \\ 0 & 0 & -2 \end{bmatrix}$$

At this point it is clear that the system has no free variables (and that the coefficient matrix has a pivot in every column), so theorem 5 gives us that the homogeneous system has only the trivial solution.

4b: Using the definition of linear independence, we need to determine the nature of the solutions of

$$a \sin^2 t + b \cos^2 t = 0 \tag{7.5.14}$$

where 0 is the *zero function*, not the number zero. This means the equation has to be true for all values of t! Attempting to solve the equation for b:

$$\begin{aligned} b \cos^2 t &= -a \sin^2 t \\ b &= -a \frac{\sin^2 t}{\cos^2 t} \\ b &= -a \tan^2 t \end{aligned}$$

This last equation is true for all values of t (for which it is defined) only if $a = 0$, which forces $b = 0$. In other words, the only solution is $a = b = 0$. Since the only solution of (7.5.14) is the trivial solution, $\sin^2 t$ and $\cos^2 t$ are linearly indpendent.

5c: The double angle formula, $\cos(2t) = \cos^2 t - \sin^2 t$ suggests that

$$\sin^2 t - \cos^2 t + \cos(2t) = 0$$

for all t. Thereby we have produced a nontrivial linear combination of $\sin^2 t$, $\cos^2 t$, and $\cos(2t)$ that sums to the *zero function*, proving that $\sin^2 t$, $\cos^2 t$, and $\cos(2t)$ are linearly dependent.

6g: The question of whether the columns form a linearly independent set is a rewording of the question of whether the columns are linearly independent. By theorem 5, it is equivalent to determine whether the matrix has a pivot position in every column. If it does, part (iv) of the theorem is true, and therefore part (i) is true. If it does not, part (iv) of the theorem is false, and therefore part (i) is false.

Since the matrix has 4 rows, it can have at most 4 pivot positions. Since it has 6 columns, this means there is certainly a column (at least two, actually) without a pivot position. Therefore part (iv) of theorem 5 is false. Equivalently, part (i) is false and the columns of the matrix are linearly dependent.

7b: The question can be rephrased as a question about a linear system. By definition, the linear independence of the set hinges of the nature of the solutions of

$$a\begin{bmatrix} 1 & 8 & -11 \end{bmatrix} + b\begin{bmatrix} 9 & 4 & -7 \end{bmatrix} + c\begin{bmatrix} 4 & -2 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}. \tag{7.5.15}$$

This equation can be rewritten using algebra:

$$\begin{aligned} \begin{bmatrix} a & 8a & -11a \end{bmatrix} + \begin{bmatrix} 9b & 4b & -7b \end{bmatrix} + \begin{bmatrix} 4c & -2c & 2c \end{bmatrix} &= \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \\ \begin{bmatrix} a+9b+4c & 8a+4b-2c & -11a-7b+2c \end{bmatrix} &= \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

The only way for this last equation to be true is if a, b, c solve the linear system

$$\begin{array}{rcrcrcl} a & + & 9b & + & 4c & = & 0 \\ 8a & + & 4b & - & 2c & = & 0 \\ -11a & - & 7b & + & 2c & = & 0 \end{array}. \tag{7.5.16}$$

By row reduction,

$$\begin{bmatrix} 1 & 9 & 4 & 0 \\ 8 & 4 & -2 & 0 \\ -11 & -7 & 2 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 9 & 4 & 0 \\ 0 & -68 & -34 & 0 \\ 0 & -92 & 46 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 9 & 4 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 2 & -1 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 9 & 4 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & -2 & 0 \end{bmatrix}$$

At this point, it is clear that the system has no free variables. Therefore part (vii) of theorem 5 is true. The equivalent part (iii) is therefore true, so (7.5.16) has only the trivial solution. In turn, the trivial solution is the only solution of (7.5.15), so the set is linearly independent.

18: To show that two statements are true requires showing that if one of them is true, so is the other, and vice versa, so there are two things to show. (i) If (a) is true, then (b) is true; and (ii) if (b) is true, then (a) is true.

(a) $\Rightarrow$ (b): **[We start by assuming (a) is true and showing that (b) follows logically.]** Suppose $x = 8$. Then x is a perfect cube since $2^3 = 8$. Of course $6 < 8 < 20$ so $6 < x < 20$. Hence x is a perfect cube between 6 and 20.

(b) $\Rightarrow$ (a): **[We now assume (b) is true and show that (a) follows logically.]** Suppose x is a perfect cube between 6 and 20. Because x is a perfect cube, it must be one of the numbers $1, 8, 27$ or higher. The only one of those numbers between 6 and 20 is 8, so x must be 8.

Section 3.4

5d: According to theorem 6 it is equivalent to show that the coefficient matrix has a pivot position in every row. This is done by row reduction:

$$\begin{bmatrix} 6 & 5 & 1 & 5 \\ 3 & 0 & -4 & 7 \\ -1 & 1 & 1 & -4 \end{bmatrix} \to \begin{bmatrix} -1 & 1 & 1 & -4 \\ 3 & 0 & -4 & 7 \\ 6 & 5 & 1 & 5 \end{bmatrix} \to \begin{bmatrix} -1 & 1 & 1 & -4 \\ 0 & 3 & -1 & -5 \\ 0 & 11 & 7 & -19 \end{bmatrix}$$
$$\to \begin{bmatrix} -1 & 1 & 1 & -4 \\ 0 & 33 & -11 & -55 \\ 0 & -33 & -21 & 57 \end{bmatrix} \to \begin{bmatrix} -1 & 1 & 1 & -4 \\ 0 & 33 & -11 & -55 \\ 0 & 0 & -32 & 2 \end{bmatrix}$$

At this point it is clear there is a pivot in every row. Therefore the system has only the trivial solution.

6e: According to theorem 6 it is equivalent to answer the question "Does the matrix have a pivot position in every row?" This is determined by row reduction:

$$\begin{bmatrix} -18 & 1 & -1 & 7 \\ -6 & 2 & 6 & 3 \\ 12 & -5 & 0 & 4 \end{bmatrix} \to \begin{bmatrix} -6 & 2 & 6 & 3 \\ -18 & 1 & -1 & 7 \\ 12 & -5 & 0 & 4 \end{bmatrix} \to \begin{bmatrix} -6 & 2 & 6 & 3 \\ 0 & -5 & -19 & -2 \\ 0 & -1 & 12 & 10 \end{bmatrix}$$
$$\to \begin{bmatrix} -6 & 2 & 6 & 3 \\ 0 & -1 & 12 & 10 \\ 0 & -5 & -19 & -2 \end{bmatrix} \to \begin{bmatrix} -6 & 2 & 6 & 3 \\ 0 & -1 & 12 & 10 \\ 0 & 0 & -79 & -52 \end{bmatrix}$$

At this point it is clear there is a pivot in every row. Therefore the rows form a lineraly independent set.

6g: According to theorem 6 it is equivalent to answer the question "Does the matrix have a pivot position in every row?" Since the matrix has 5 columns, it can have at most 5 pivot positions. However the matrix has 6 rows, so cannot have a pivot in every row. Consequently the rows do not form a linearly independent set.

17: What makes any matrix M upper triangular is that $M_{k,\ell} = 0$ whenever $k > \ell$. That is, entries whose row number is greater than their column number are zero. After deleting the first row and some column of U, as shown below,

$$\begin{bmatrix} \boxed{\star} & \boxed{\star} & \boxed{\star} & \boxed{\star} & \boxed{\cdots} & \boxed{\star} \\ 0 & \star & \star & \boxed{\star} & \cdots & \star \\ 0 & 0 & \star & \boxed{\star} & \cdots & \star \\ 0 & 0 & 0 & \boxed{\star} & \cdots & \star \\ \vdots & \vdots & \vdots & \boxed{\vdots} & \ddots & \star \\ 0 & 0 & 0 & \boxed{0} & \cdots & \star \end{bmatrix}$$

there are two distinct regions of entries. Entries to the left of the deleted column have the same column index in $U_{\backslash i,j}$ as they do in U. Columns to the right of the deleted column have a column index one less than they do in U. All entries in $U_{\backslash 1,j}$ have a row index one less than they do in U. In symbols [and **the start of the proof**],

$$(U_{\backslash 1,j})_{k,\ell} = \begin{cases} U_{k+1,\ell} & \text{if } \ell < j \\ U_{k+1,\ell+1} & \text{if } \ell \geq j \end{cases}.$$

If $k > \ell$, then $k+1 > \ell$ and $k+1 > \ell+1$, so $U_{k+1,\ell} = U_{k+1,\ell+1} = 0$. Hence $(U_{\backslash 1,j})_{k,\ell} = 0$ whenever $k > \ell$.

18: As long as $j > 1$, $(U_{\backslash 1,j})_{1,1} = U_{2,1} = 0$.

19: ***[Commentary that is not strictly part of the proof wil be inserted in square brackets and bold italicized.]*** If U is a 1×1 matrix, it is upper triangular and $\det([U_{1,1}]) = U_{1,1}$. ***[This establishes part (i) of the proof. The claim is true for the particular value*** $n = 1$***.]*** Now assume that $\det U = U_{1,1}U_{2,2}\cdots U_{k,k}$ for some (arbitrary) value k greater than or equal to one and arbitrary upper triangular $k \times k$ matrix U. ***[That is, if*** U ***is an upper triangular*** $k \times k$ ***matrix and*** $k \geq 1$***, then the proposition is true. To complete the proof, we must use this information to***

***prove that if** U **is a** $(k+1)\times(k+1)$ **upper triangular matrix then** $\det U = U_{1,1}U_{2,2}\cdots U_{k+1,k+1}$.**]* Additionally, suppose U is a $(k+1)\times(k+1)$ upper triangular matrix. By definition,

$$\det U = (-1)^{1+1}U_{1,1}\det U_{\backslash 1,1} + (-1)^{1+2}U_{1,2}\det U_{\backslash 1,2}\cdots + (-1)^{1+3}U_{1,3}\det U_{\backslash 1,3}.$$

Since $U_{\backslash 1,j}$ has a zero on its diagonal whenever $j > 1$, all terms except the first are zero. Therefore,

$$\det U = (-1)^{1+1}U_{1,1}\det U_{\backslash 1,1}. \tag{7.5.17}$$

Since $U_{\backslash 1,1}$ is a $k\times k$ matrix, its determinant is the product of the entries on its diagonal ***[this is the inductive hypothesis]***, so $\det U_{\backslash 1,1} = U_{2,2}U_{3,3}\cdots U_{k+1,k+1}$. Substituting this expresion into (7.5.17), we have $\det U = U_{1,1}U_{2,2}U_{3,3}\cdots U_{k+1,k+1}$, and the proof is complete.

Section 3.5

1a: It is advantageous to expand along rows and columns with many zeros since corresponding cofactors will not need to be computed. They will be multiplied by the zero entry in the computation of the determinant anyway. In this case, the fourth row has three zeros:

$$\begin{vmatrix} 0 & -2 & 0 & 9 \\ -4 & 0 & 1 & 0 \\ 4 & -9 & 0 & -2 \\ 0 & 0 & 0 & 4 \end{vmatrix} = 4(1)\begin{vmatrix} 0 & -2 & 0 \\ -4 & 0 & 1 \\ 4 & -9 & 0 \end{vmatrix} = 4(-2)(-1)\begin{vmatrix} -4 & 1 \\ 4 & 0 \end{vmatrix} = 8(-4) = -32$$

The 3×3 determinant is expanded along the first row since it contains two zeros. The 2×2 determinant, at this point, is probably as simple as having memorized the pattern (upper left times lower right minus upper right times lower left).

2c: The determinant of any triangular matrix is the product of the entries on its main diagonal. In this case, $2\cdot 8\cdot 4\cdot 7 = 448$

3: (a) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is a swap matrix, so the determinant of the product is -1 times the determinant of the given matrix (see justification on page 102): $(-1)(32) = -32$.

(b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{128} \end{bmatrix}$ is a scale matrix, so the determinant of the product is the scale factor, $\frac{1}{128}$, times the determinant of the given matrix (see justification on page 102): $\frac{1}{128}(32) = \frac{1}{4}$.

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{7}{9} \\ 0 & 0 & 1 \end{bmatrix}$ is a replacement matrix, so the determinant of the product is the same as the determinant of the given matrix (see justification on page 102): 32.

(d) The matrices $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 18 & 0 & 1 \end{bmatrix}$ are a scale by 3 and a replacement. Their effect on the determinant of the product is, altogether, to multiply it by 3: $3(32) = 96$.

Section 3.6

1: The determinant of the given matrix is 120. (a) Row replacements do not change the determinant of a matrix, so the original matrix must have had determinant 120 as well. (b) Row replacements do not change the determinant of a matrix, and row swaps change the sign of the determinant. Therefore $(\det A)(-1)^3 = 120$ so $\det A = -120$. (c) Row replacements do not change the determinant of a matrix, and row scaling multiplies the determinant by the scale factor. Therefore $(\det A)(6)(5) = 120$ so $\det A = 4$. Row replacements do not change the determinant of a matrix, row scaling multiplies the determinant by the scale factor, and row swaps change the sign of the determinant. Therefore $(\det A)(10)(-1)^2 = 120$ so $\det A = 12$.

2b: The determinant of the given matrix is the product of the entries on the main diagonal (since the matrix is triangular): $(-2)(-1)(5)(6) = 60$. Row replacements do not change the determinant of a matrix, and row swaps change the sign of the determinant. Therefore $(\det A)(-1)^k = 60$ where k is the number of row swaps. Since $(-1)^k$ equals 1 or -1, $\det A = \pm 60$.

5b: Row reducing:

$$\begin{bmatrix} -12 & 12 \\ 14 & 6 \end{bmatrix} \xrightarrow{-\frac{1}{12}M_{1,:}\to M_{1,:}} \begin{bmatrix} 1 & 1 \\ 14 & 6 \end{bmatrix} \xrightarrow{-14M_{1,:}+M_{2,:}\to M_{2,:}} \begin{bmatrix} 1 & 1 \\ 0 & -8 \end{bmatrix}$$

The determinant of a triangular matrix is the product of the entries of the diagonal, so $\det\begin{pmatrix} 1 & 1 \\ 0 & -8 \end{pmatrix} = -8$. The triangular matrix was gotten by scaling with factor $-\frac{1}{12}$ and one row replacement. Hence the determinant of M satisfies $(\det M)\left(-\frac{1}{12}\right) = -8$ so $\det M = 96$.

4f: Row reducing:

$$\begin{bmatrix} -11 & -15 & 4 \\ 8 & 9 & -4 \\ -3 & -3 & 2 \end{bmatrix} \xrightarrow[-11M_{3,:}\to M_{3,:}]{11M_{2,:}\to M_{2,:}} \begin{bmatrix} -11 & -15 & 4 \\ 88 & 99 & -44 \\ 33 & 33 & -22 \end{bmatrix} \xrightarrow[3M_{1,:}+M_{3,:}\to M_{3,:}]{8M_{1,:}+M_{2,:}\to M_{2,:}} \begin{bmatrix} -11 & -15 & 4 \\ 0 & -21 & -12 \\ 0 & -12 & -10 \end{bmatrix}$$

$$\xrightarrow[\frac{7}{2}M_{3,:}\to M_{3,:}]{-\frac{1}{3}M_{2,:}\to M_{2,:}} \begin{bmatrix} -11 & -15 & 4 \\ 0 & 7 & 4 \\ 0 & -42 & -35 \end{bmatrix} \xrightarrow{6M_{2,:}+M_{3,:}\to M_{3,:}} \begin{bmatrix} -11 & -15 & 4 \\ 0 & 7 & 4 \\ 0 & 0 & -11 \end{bmatrix}$$

The determinant of a triangular matrix is the product of the entries of the diagonal, so

$$\det\begin{pmatrix} -11 & -15 & 4 \\ 0 & 7 & 4 \\ 0 & 0 & -11 \end{pmatrix} = 7(11)^2.$$

The triangular matrix was gotten by scaling with factors $11, -11, -\frac{1}{3}, \frac{7}{2}$ and three row replacements. Hence the determinant of M satisfies $(\det M)(11)(-11)\left(-\frac{1}{3}\right)\left(\frac{7}{2}\right) = 7(11)^2$ so

$$\det M = \frac{7(11)^2(2)(3)}{11(-11)(-7)} = 6.$$

4g: Row reducing:

$$\begin{bmatrix} 3 & 90 & -308 & -6 \\ -3 & -140 & 484 & 10 \\ 6 & 210 & -737 & -16 \\ 3 & 70 & -231 & -4 \end{bmatrix} \xrightarrow[\substack{-2M_{1,:}+M_{3,:}\to M_{3,:} \\ -M_{1,:}+M_{4,:}\to M_{4,:}}]{M_{1,:}+M_{2,:}\to M_{2,:}} \begin{bmatrix} 3 & 90 & -308 & -6 \\ 0 & -50 & 176 & 4 \\ 0 & 30 & -121 & -4 \\ 0 & -20 & 77 & 2 \end{bmatrix} \xrightarrow[-5M_{4,:}\to M_{4,:}]{5M_{3,:}\to M_{3,:}}$$

$$\begin{bmatrix} 3 & 90 & -308 & -6 \\ 0 & -50 & 176 & 4 \\ 0 & 150 & -605 & -20 \\ 0 & 100 & -385 & -10 \end{bmatrix} \xrightarrow[2M_{2,:}+M_{4,:}\to M_{4,:}]{3M_{2,:}+M_{3,:}\to M_{3,:}} \begin{bmatrix} 3 & 90 & -308 & -6 \\ 0 & -50 & 176 & 4 \\ 0 & 0 & -77 & -8 \\ 0 & 0 & -33 & -2 \end{bmatrix} \xrightarrow{-7M_{4,:}\to M_{4,:}}$$

$$\begin{bmatrix} 3 & 90 & -308 & -6 \\ 0 & -50 & 176 & 4 \\ 0 & 0 & -77 & -8 \\ 0 & 0 & 231 & 14 \end{bmatrix} \xrightarrow{3M_{3,:}+M_{4,:}\to M_{4,:}} \begin{bmatrix} 3 & 90 & -308 & -6 \\ 0 & -50 & 176 & 4 \\ 0 & 0 & -77 & -8 \\ 0 & 0 & 0 & -10 \end{bmatrix}$$

The determinant of a triangular matrix is the product of the entries of the diagonal, so

$$\det\begin{pmatrix} 3 & 90 & -308 & -6 \\ 0 & -50 & 176 & 4 \\ 0 & 0 & -77 & -8 \\ 0 & 0 & 0 & -10 \end{pmatrix} = -(3)(50)(77)(10).$$

The triangular matrix was gotten by scaling with factors $5, -5, -7$, and three row replacements. Hence the determinant of M satisfies $(\det M)(5)(-5)(-7) = -3(50)(77)(10)$ so

$$\det M = \frac{-3(50)(77)(10)}{(5)(-5)(-7)} = -3(10)(-11)(-2) = -660.$$

5b: Eliminating the 4 by row replacement and then expanding along the first column:

$$\begin{bmatrix} -1 & 5 & -1 \\ 0 & 6 & -1 \\ 4 & -10 & 2 \end{bmatrix} \xrightarrow{4M_{1,:}+M_{3,:}\to M_{3,:}} \begin{bmatrix} -1 & 5 & -1 \\ 0 & 6 & -1 \\ 0 & 10 & -2 \end{bmatrix}$$

and

$$\begin{vmatrix} -1 & 5 & -1 \\ 0 & 6 & -1 \\ 0 & 10 & -2 \end{vmatrix} = (-1)\begin{vmatrix} 6 & -1 \\ 10 & -2 \end{vmatrix} = (-1)(-12+10) = 2.$$

Since the original matrix was only changed by row replacement, its determinant and the determinant of the resulting matrix are equal. Hence

$$\begin{vmatrix} -1 & 5 & -1 \\ 0 & 6 & -1 \\ 4 & -10 & 2 \end{vmatrix} = 2.$$

6e: Theorem 7 gives a number of conditions equivalent to invertibility of a square matrix M. Among them are (vi) M has a pivot position in every column; (vii) M has a pivot position in every row; and (xiii) $\det M \neq 0$. If we can show any one of these, we will know that the given matrix is invertible. Row reducing:

$$\begin{bmatrix} 1 & 1 & 3 \\ 4 & 7 & 11 \\ 0 & 3 & 20 \end{bmatrix} \xrightarrow{-4M_{1,:}+M_{2,:}\to M_{2,:}} \begin{bmatrix} 1 & 1 & 3 \\ 0 & 3 & -1 \\ 0 & 3 & 20 \end{bmatrix} \xrightarrow{-M_{2,:}+M_{3,:}\to M_{3,:}} \begin{bmatrix} 1 & 1 & 3 \\ 0 & 3 & -1 \\ 0 & 0 & 21 \end{bmatrix}$$

At this point, we can see all of (vi), (vii), and (xiii) are true. We may pick any one of them to explain why the given matrix is invertible. For example, the given matrix has nonzero determinant and is therefore invertible.

Section 3.7

1c: Augmenting the identity matrix and reducing:

$$\begin{bmatrix} 9 & -7 & 1 & 0 \\ 4 & 7 & 0 & 1 \end{bmatrix} \xrightarrow{-2M_{2,:}+M_{1,:}\to M_{1,:}} \begin{bmatrix} 1 & -21 & 1 & -2 \\ 4 & 7 & 0 & 1 \end{bmatrix} \xrightarrow{-4M_{1,:}+M_{2,:}\to M_{2,:}} \begin{bmatrix} 1 & -21 & 1 & -2 \\ 0 & 91 & -4 & 9 \end{bmatrix}$$

$$\xrightarrow{\frac{1}{91}M_{2,:}\to M_{2,:}} \begin{bmatrix} 1 & -21 & 1 & -2 \\ 0 & 1 & -\frac{4}{91} & \frac{9}{91} \end{bmatrix} \xrightarrow{21M_{2,:}+M_{1,:}\to M_{1,:}} \begin{bmatrix} 1 & 0 & \frac{7}{91} & \frac{7}{91} \\ 0 & 1 & -\frac{4}{91} & \frac{9}{91} \end{bmatrix}$$

1d: Augmenting the identity matrix and reducing:

$$\begin{bmatrix} 16 & -3 & -2 & 1 & 0 & 0 \\ -8 & 4 & -2 & 0 & 1 & 0 \\ -8 & 1 & 2 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{M_{1,:}\leftrightarrow M_{3,:}} \begin{bmatrix} -8 & 1 & 2 & 0 & 0 & 1 \\ -8 & 4 & -2 & 0 & 1 & 0 \\ 16 & -3 & -2 & 1 & 0 & 0 \end{bmatrix} \xrightarrow[2M_{1,:}+M_{2,:}\to M_{2,:}]{-M_{1,:}+M_{2,:}\to M_{2,:}}$$

$$\left[\begin{array}{cccccc} -8 & 1 & 2 & 0 & 0 & 1 \\ 0 & 3 & -4 & 0 & 1 & -1 \\ 0 & -1 & 2 & 1 & 0 & 2 \end{array}\right] \xrightarrow{M_{2,:}\leftrightarrow M_{3,:}} \left[\begin{array}{cccccc} -8 & 1 & 2 & 0 & 0 & 1 \\ 0 & -1 & 2 & 1 & 0 & 2 \\ 0 & 3 & -4 & 0 & 1 & -1 \end{array}\right] \xrightarrow{3M_{2,:}+M_{3,:}\leftrightarrow M_{3,:}}$$

$$\left[\begin{array}{cccccc} -8 & 1 & 2 & 0 & 0 & 1 \\ 0 & -1 & 2 & 1 & 0 & 2 \\ 0 & 0 & 2 & 3 & 1 & 5 \end{array}\right] \underset{-M_{3,:}+M_{2,:}\to M_{2,:}}{\xrightarrow{-M_{3,:}+M_{1,:}\to M_{1,:}}} \left[\begin{array}{cccccc} -8 & 1 & 0 & -3 & -1 & -4 \\ 0 & -1 & 0 & -2 & -1 & -3 \\ 0 & 0 & 2 & 3 & 1 & 5 \end{array}\right] \underset{-M_{2,:}\to M_{2,:}}{\xrightarrow{M_{2,:}+M_{1,:}\to M_{1,:}}}$$

$$\left[\begin{array}{cccccc} -8 & 0 & 0 & -5 & -2 & -7 \\ 0 & 1 & 0 & 2 & 1 & 3 \\ 0 & 0 & 2 & 3 & 1 & 5 \end{array}\right] \underset{\frac{1}{2}M_{3,:}\to M_{3,:}}{\xrightarrow{-\frac{1}{8}M_{1,:}\to M_{1,:}}} \left[\begin{array}{cccccc} 1 & 0 & 0 & \frac{5}{8} & \frac{1}{4} & \frac{7}{8} \\ 0 & 1 & 0 & 2 & 1 & 3 \\ 0 & 0 & 1 & \frac{3}{2} & \frac{1}{2} & \frac{5}{2} \end{array}\right]$$

2: Three facts are applied.

1. [section 3.7] The determinant of a product is the product of the determinant [$\det(AB) = (\det A)(\det B)$].
2. [section 3.7] The determinant of an inverse is the reciprocal of the determinant [$\det(A^{-1}) = \frac{1}{\det A}$].
3. [section 3.5] The determinant of a transpose equals the determinant [$\det(A^T) = \det A$].

(a) $\det(MR^T) = (\det M)(\det R^T) = (\det M)(\det R) = 2 \cdot \frac{1}{3} = \frac{2}{3}$ using facts 1 and 3.
(b) $\det(M^{-1}R) = (\det M^{-1})(\det R) = \frac{1}{\det M} \det R = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$ using facts 1 and 2.
(c) $\det(MR^{-1})^T = \det(MR^{-1}) = (\det M)(\det R^{-1}) = \det M \frac{1}{\det R} = 2 \cdot 3 = 6$ using facts 1,2 and 3.

Section 4.1

1b: Without any information to whittle down the list of properties to verify, all 10 of the properties from the definition of vector space must be verified. Let $\mathbf{f}, \mathbf{g}, \mathbf{h}$ be in V (letters commonly associated with functions are used to emphasize that these vectors are in fact functions) and let s, t be scalars. Two functions are equal if they take the same value (give the same output) on every point in the domain.

1. $\mathbf{f} + \mathbf{g}$ means $\mathbf{f}(x) + \mathbf{g}(x)$ and is a function defined on $[0, 1]$ since $\mathbf{f}(x)$ and $\mathbf{g}(x)$ are. Therefore $\mathbf{f} + \mathbf{g}$ is in V.
2. $\mathbf{f}(x) + \mathbf{g}(x) = \mathbf{g}(x) + \mathbf{f}(x)$ for every x in $[0, 1]$ since addition of real numbers is commutative. Therefore $\mathbf{f} + \mathbf{g} = \mathbf{g} + \mathbf{f}$.
3. $\mathbf{f}(x) + (\mathbf{g}(x) + \mathbf{h}(x)) = (\mathbf{g}(x) + \mathbf{f}(x)) + \mathbf{h}(x)$ for every x in $[0, 1]$ since addition of real numbers is associative. Therefore $\mathbf{f} + (\mathbf{g} + \mathbf{h}) = (\mathbf{g} + \mathbf{f}) + \mathbf{h}$.
4. The function $\mathbf{z}(x) = 0$ (whose graph is a horizontal line on the x-axis) is in V and $\mathbf{z}(x) + \mathbf{f}(x) = \mathbf{f}(x)$ for all x in $[0, 1]$. Therefore $\mathbf{z}$ is an additive identity in V.
5. The function $-\mathbf{f}$ is in V since it is a function on $[0, 1]$ and has the property that $\mathbf{f}(x) + (-\mathbf{f}(x)) = 0$ for all x in $[0, 1]$. Therefore $\mathbf{f} + (-\mathbf{f}) = \mathbf{z}$ so every element of V has an additive inverse.
6. $s\mathbf{f}(x)$ is a function on $[0, 1]$ because $\mathbf{f}(x)$ is. Therefore $s\mathbf{f}$ is in V.
7. $1\mathbf{f}(x) = \mathbf{f}(x)$ for all x in $[0, 1]$. Therefore $1\mathbf{f} = \mathbf{f}$.
8. $s(\mathbf{f}(x) + \mathbf{g}(x)) = s\mathbf{f}(x) + s\mathbf{g}(x)$ for every x in $[0, 1]$ by the distributive property of real numbers. Therefore $s(\mathbf{f} + \mathbf{g}) = s\mathbf{f} + s\mathbf{g}$.
9. $(s + t)\mathbf{f}(x) = s\mathbf{f}(x) + t\mathbf{f}(x)$ for every x in $[0, 1]$ by the distributive property of real numbers. Therefore $(s + t)\mathbf{f} = s\mathbf{f} + t\mathbf{f}$.
10. $s(t\mathbf{f}(x)) = (st)\mathbf{f}(x)$ for every x in $[0, 1]$ because multiplication of real numbers is associative. Therefore $s(t\mathbf{f}) = (st)\mathbf{f}$.

2a: [To verify that S is a subspace, three properties need to be shown—that the zero vector is in S; that S is closed under addition; and that S is closed under scalar multiplication.]

In $\mathbb{R}^2$, $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is in S since it is on the x-axis (the y-coordinate is zero). **[This shows that the zero vector is in S.]** For any $\mathbf{u} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ in S $y_1 = 0$, so $\mathbf{u} = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$. It follows that $s\mathbf{u} = \begin{bmatrix} sx_1 \\ 0 \end{bmatrix}$, which is in S since it also lies on the x-axis (has y-coordinate zero). **[This shows that S is closed under scalar multiplication.]** Similarly, for any $\mathbf{v} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$ in S $y_2 = 0$. Therefore $\mathbf{u} + \mathbf{v} = \begin{bmatrix} x_1 \\ 0 \end{bmatrix} + \begin{bmatrix} x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ 0 \end{bmatrix}$, which is in S since it also lies on the x-axis (has y-coordinate zero). **[This shows that S is closed under addition.]**

2d: [To verify that S is a subspace, three properties need to be shown—that the zero vector is in S; that S is closed under addition; and that S is closed under scalar multiplication.]

In $\mathbb{P}_3(\mathbb{R})$, $\mathbf{0} = (z(x) = 0)$ and $z(x) = 0$ is in S since it is a polynomial of degree less than three with roots at 3 and 18. **[This shows that the zero vector is in S.]** For any $\mathbf{p}$ in S $\mathbf{p}(3) = \mathbf{p}(18) = 0$ and $\mathbf{p}$ is a polynomial of degree three or less. Since multiplying a polynomial by a scalar does not change its degree, $(s\mathbf{p})$ is a polynomial of degree three or less for any scalar s. Moreover $(s\mathbf{p})(3) = s\,(\mathbf{p}(3)) = s \cdot 0 = 0$ and $(s\mathbf{p})(18) = s\,(\mathbf{p}(18)) = s \cdot 0 = 0$ so $s\mathbf{p}$ has roots at 3 and 18. **[This shows that S is closed under scalar multiplication.]** Similarly, for any $\mathbf{q}$ in S $\mathbf{q}(3) = \mathbf{q}(18) = 0$ and $\mathbf{q}$ is a polynomial of degree three or less. Since the sum of two polynomials has degree no higher than the one with highest degree $\mathbf{p} + \mathbf{q}$ is a polynomial of degree at most 3. Moreover $(\mathbf{p} + \mathbf{q})(3) = \mathbf{p}(3) + \mathbf{q}(3) = 0 + 0 = 0$ and $(\mathbf{p} + \mathbf{q})(18) = \mathbf{p}(18) + \mathbf{q}(18) = 0 + 0 = 0$ so $\mathbf{p} + \mathbf{q}$ has roots at 3 and 18. **[This shows that S is closed under addition.]**

3a: [To show that S is not a subspace, one of the three subspace properties needs to be shown false—that the zero vector is in S; that S is closed under addition; and that S is closed under scalar multiplication. The easiest way to do this is often by counterexample (an example showing that one of the properties does not hold in all instances).] $\begin{bmatrix} 3 \\ 14 \end{bmatrix}$ is in S but $-2\begin{bmatrix} 3 \\ 14 \end{bmatrix} = \begin{bmatrix} -6 \\ -28 \end{bmatrix}$ is not, so S is not closed under multiplication. **Note: S contains the zero vector and is closed under addition, so the only property for which a counterexample can be found is closure under scalar multiplication. There are many counterexamples available. Only one is needed.**

3d: [To show that S is not a subspace, one of the three subspace properties needs to be shown false—that the zero vector is in S; that S is closed under addition; and that S is closed under scalar multiplication. The easiest way to do this is often by counterexample (an example showing that one of the properties does not hold in all instances).] $\mathbf{0} = (z(x) = 0)$ is not in S since $z(0) = 0$ (has y-intercept 0, not 3). Consequently the zero vector is not in S. **Note: S is not closed under addition nor closed under scalar multiplication either. There are many exceptions to the properties available. Only one is needed.**

4a: By definition,

$$\operatorname{span}S = \left\{ r\begin{bmatrix} 5 \\ 1 \end{bmatrix} : r \text{ in } \mathbb{R} \right\}$$

so span$S$ contains the vector $\begin{bmatrix} 5 \\ 1 \end{bmatrix}$ and all multiples thereof, including $0\begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Each multiple points in the same or opposite direction, and there is a multiple for every magnitude. Therefore all the points on the line containing $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 5 \\ 1 \end{bmatrix}$ are in S and nothing more. In summary, S is the line passing through $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 5 \\ 1 \end{bmatrix}$.

4c: By definition,

$$\operatorname{span}S = \left\{ r\begin{bmatrix} 5 \\ 1 \end{bmatrix} + s\begin{bmatrix} 2 \\ 0 \end{bmatrix} : r, s \text{ in } \mathbb{R} \right\}.$$

Equivalently, $\operatorname{span}S = \left\{ \begin{bmatrix} 5 & 2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} r \\ s \end{bmatrix} : r, s \text{ in } \mathbb{R} \right\}$ since a matrix times a vector is a linear combination of the columns of the matrix. Since $\begin{bmatrix} 5 & 2 \\ 1 & 0 \end{bmatrix}$ has nonzero determinant $(5 \cdot 0 - 2 \cdot 1)$ theorem 7 assures that $\begin{bmatrix} 5 & 2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} r \\ s \end{bmatrix} = \mathbf{b}$ has exactly one solution for every $\mathbf{b}$ in $\mathbb{R}^2$. In other words, every vector in $\mathbb{R}^2$ is in the span of S, so $\operatorname{span}S = \mathbb{R}^2$.

5f: By definition,

$$\operatorname{span} S = \left\{ a\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + c\begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$$

so $\operatorname{span} S = \left\{ \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & b \end{bmatrix} - \begin{bmatrix} 0 & c \\ 0 & 0 \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\} = \left\{ \begin{bmatrix} a & -c \\ 0 & b \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$. Since a, b, c are arbitrary real numbers, this is equivalent to $\left\{ \begin{bmatrix} r_1 & r_2 \\ 0 & r_3 \end{bmatrix} : r_1, r_2, r_3 \text{ in } \mathbb{R} \right\}$, which is the set of all 2×2 matrices with a zero in the 2,1-entry.

8: Since $\begin{bmatrix} 3t \\ -t \\ 5t \end{bmatrix} = t\begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$, $S = \left\{ t\begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix} ; t \text{ in } \mathbb{R} \right\}$, and S contains all multiples of $\begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$, or in other words all linear combinations of vectors in $\left\{ \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix} \right\}$. Hence the desired set is $\left\{ \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix} \right\}$.

9: Since $\begin{bmatrix} 3t - 2s \\ -t + s \\ 5t + s \end{bmatrix} = t\begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix} + s\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $S = \left\{ t\begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix} + s\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} ; t, s \text{ in } \mathbb{R} \right\}$. This is the set of all linear combinations of $\begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$ and $\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, the very definition of $\operatorname{span}\left\{ \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \right\}$. Hence the desired set is $\left\{ \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \right\}$.

12b: By definition, $\operatorname{span} S = \left\{ x\begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} + y\begin{bmatrix} 10 \\ 2 \\ -6 \end{bmatrix} : x, y \text{ in } \mathbb{R} \right\}$ so the question is whether $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ is in this set. In other words, can we solve the equation $x\begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} + y\begin{bmatrix} 10 \\ 2 \\ -6 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$? Written in matrix form, that is $\begin{bmatrix} 7 & 10 \\ 8 & 2 \\ 9 & -6 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and whether there is a solution can be determined by reducing the augmented matrix:

$$\begin{bmatrix} 7 & 10 & 1 \\ 8 & 2 & 2 \\ 9 & -6 & 3 \end{bmatrix} \xrightarrow{\text{reduces to}} \begin{bmatrix} 1 & 58 & -5 \\ 0 & 66 & -6 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence the system is consistent. There is a solution, and therefore $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ is in $\operatorname{span} S$.

13c: By definition, $\operatorname{span}\left\{\sin^2\theta, \cos^2\theta\right\} = \left\{a\sin^2\theta + b\cos^2\theta : a, b \text{ in } \mathbb{R}\right\}$ so the question is whether $\cos(2\theta)$ is in this set. Since $\cos(2\theta) = \cos^2\theta - \sin^2\theta = 1\cos^2\theta + (-1)\sin^2\theta$ [this is a standard trig identity—double angle formula], $\cos(2\theta)$ is in $\operatorname{span}\left\{\sin^2\theta, \cos^2\theta\right\}$.

14c: Since $\mathbf{v} = \langle 78, 81, 84, 87, 90, \ldots \rangle$ can be written as

$$3\langle 1, 2, 3, 4, 5, \ldots \rangle + 75\langle 1, 1, 1, 1, 1, \ldots \rangle$$

$\mathbf{v}$ is a linear combination of the vectors in the given set. Since the span includes all linear combinations of its elements, it includes this one. Therefore the answer is yes.

15a: Implied in the question is that we are to treat the columns of the matrix as individual vectors. Therefore we are to determine whether $\begin{bmatrix} -240 \\ -406 \\ -416 \end{bmatrix}$ is in

$$\left\{ x\begin{bmatrix} -499 \\ -425 \\ 306 \end{bmatrix} + y\begin{bmatrix} -288 \\ -161 \\ -348 \end{bmatrix} + z\begin{bmatrix} 232 \\ 125 \\ 141 \end{bmatrix} : x, y, z \text{ in } \mathbb{R} \right\}.$$

But this set is exactly $\left\{ \begin{bmatrix} -499 & -288 & 232 \\ -425 & -161 & 125 \\ 306 & -348 & 141 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x, y, z \text{ in } \mathbb{R} \right\}$ so we need to determine whether

$$\begin{bmatrix} -499 & -288 & 232 \\ -425 & -161 & 125 \\ 306 & -348 & 141 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -240 \\ -406 \\ -416 \end{bmatrix} \tag{7.5.18}$$

has a solution, which can be done be reducing the augmented matrix $\begin{bmatrix} -499 & -288 & 232 & -240 \\ -425 & -161 & 125 & -406 \\ 306 & -348 & 141 & -416 \end{bmatrix}$. Adding the one line `A.rref()` to the SageMath code produces

```
[                1                0                0  1533968/1012773]
[                0                1                0   2328386/337591]
[                0                0                1 10922888/1012773]
```

for the reducded matrix, indicating that system (7.5.18) is consistent (has a solution). Yes, **b** is in the span of the columns of M.

Section 4.2

4d: (i) This is a subset of $\mathcal{M}_{2\times 2}(\mathbb{R})$. (ii) The content of the final paragraph of this section (section 4.2) is that any vector that can be written as a linear combination of the others in a set can be removed without affecting the span of the set. In this case, $\begin{bmatrix} -4 & -19 \\ 14 & -13 \end{bmatrix} + \begin{bmatrix} -19 & -2 \\ 14 & -1 \end{bmatrix} = \begin{bmatrix} -23 & -21 \\ 28 & -14 \end{bmatrix}$ so $\begin{bmatrix} -23 & -21 \\ 28 & -14 \end{bmatrix}$ is a linear combination of the others and we can remove it without affecting the span. Hence $\left\{ \begin{bmatrix} -4 & -19 \\ 14 & -13 \end{bmatrix}, \begin{bmatrix} -19 & -2 \\ 14 & -1 \end{bmatrix} \right\}$ is a subset with the same span.

5d: $\begin{bmatrix} -4 & -19 \\ 14 & -13 \end{bmatrix} = \begin{bmatrix} -23 & -21 \\ 28 & -14 \end{bmatrix} - \begin{bmatrix} -19 & -2 \\ 14 & -1 \end{bmatrix}$ so $\begin{bmatrix} -4 & -19 \\ 14 & -13 \end{bmatrix}$ can be removed from the set without affecting its span. Hence $\left\{ \begin{bmatrix} -23 & -21 \\ 28 & -14 \end{bmatrix}, \begin{bmatrix} -19 & -2 \\ 14 & -1 \end{bmatrix} \right\}$ is a (different) subset with the same span. Yes, this subset is a basis for the span because it is both spanning (as we know) and linearly independent (since it contains two elements which are not multiples of one another).

6c: Let $S = \left\{ 1 + 2t, 3 + t - 2t^2, -5 + 4t^2, -5t - 2t^2 \right\}$. Since S is a 4-element subset of $\mathbb{P}_2(\mathbb{R})$, which has dimension 3, we know that S is linearly dependent (theorem 9). Hence the equation

$$A(1 + 2t) + B(3 + t - 2t^2) + C(-5 + 4t^2) + D(-5t - 2t^2) = 0 \tag{7.5.19}$$

has nontrivial solutions. We need to find one. Expanding the lefthand side and collecting like terms:

$$A + 2At + 3B + Bt - 2Bt^2 - 5C + 4Ct^2 - 5Dt - 2Dt^2 = 0$$
$$(A + 3B - 5C) + (2A + B - 5D)t + (-2B + 4C - 2D)t^2 = 0$$

For this equation to be true (for all t, making the lefthand side equal to the zero function, the zero vector of the vector space), it must be that each coefficient is zero, giving the linear system

$$\begin{aligned} A + 3B - 5C &= 0 \\ 2A + B - 5D &= 0 \\ -2B + 4C - 2D &= 0 \end{aligned}$$

which we can solve in a number of ways. One solution is $A = -1$, $B = 2$, $C = 1$, $D = 0$. Hence

$$-1(1 + 2t) + 2(3 + t - 2t^2) + (-5 + 4t^2) + 0(-5t - 2t^2) = 0$$

and we can write (for example) $1 + 2t$ as a linear combination of the others:

$$1 + 2t = -2(3 + t - 2t^2) - (-5 + 4t^2)$$

[we could have chosen to write $(3 + t - 2t^2)$ **or** $(-5 + 4t^2)$ **as a linear combination of the others just as well]**. This means we can remove $1 + 2t$ from the set and retain its span. If the set of the three remaining vectors is linearly independent, it is a basis. If it is linearly dependent, we need to eliminate one of the vectors from $\left\{3 + t - 2t^2, -5 + 4t^2, -5t - 2t^2\right\}$. Repeating the above process without the polynomial $1 + 2t$ amounts to eliminating A from the equations: does

$$B(3 + t - 2t^2) + C(-5 + 4t^2) + D(-5t - 2t^2) = 0$$

have nontrivial solutions? We check as before by solving the linear system

$$\begin{aligned} 3B - 5C &= 0 \\ B - 5D &= 0 \\ -2B + 4C - 2D &= 0. \end{aligned}$$

One solution is $B = 5$, $C = 3$, $D = 1$. Hence

$$5(3 + t - 2t^2) + 3(-5 + 4t^2) + (-5t - 2t^2) = 0$$

and we can write (for example) $-5t - 2t^2$ as a linear combination of the others:

$$-5t - 2t^2 = -5(3 + t - 2t^2) - 3(-5 + 4t^2)$$

[we could have chosen to write $(3 + t - 2t^2)$ **or** $(-5 + 4t^2)$ **as a linear combination of the others just as well]**. This means we can remove $-5t - 2t^2$ from the set and retain its span, leaving two vectors (polynomials), $(3 + t - 2t^2)$ and $(-5 + 4t^2)$, that are not multiples of one another. Hence $\left\{3 + t - 2t^2, -5 + 4t^2\right\}$ is a linearly independent subset of S with the same span—a basis for the span of S. The dimension of the span is two.

12b: Since the dimension of $\mathbb{R}^4$ is four, a linearly independent spanning set with 4 elements will be a basis. The given set (of columns of the matrix) has four elements, so we need to determine whether it is linearly independent and spanning. Upon inspection, columns two and three are identical (making each one a linear combination of the other vectors in the set), so the set of columns is linearly dependent and does not form a basis for $\mathbb{R}^4$.

13b: Since the dimension of $\mathbb{R}^6$ is six, a linearly independent spanning set with 6 elements will be a basis. The given set (of columns of the matrix) has six elements, so we need to determine whether it is linearly independent and spanning. By theorem 7, determining linear independence is equivalent to determining whether there is a pivot position in every column—done by row reduction. Adding the line `print(M.rref())` to the provided SageCell produces the reduced matrix

```
[       1       0       0       0       0 -501/35]
[       0       1       0       0       0   273/5]
[       0       0       1       0       0   71/15]
[       0       0       0       1       0      -2]
[       0       0       0       0       1     4/3]
[       0       0       0       0       0       0]
```

revealing that the matrix does not have a pivot position in every row (or column). Hence the columns of the matrix are linearly dependent and do not form a basis for $\mathbb{R}^6$. **[Had there been a pivot position in each column, we would have needed to address the question of whether the columns were spanning—theorem 7 part (ix) provides the answer.]**

Section 4.3

4: (a) Answers will vary. Any set that contains all the ordered pairs' first components will do. For example, $\{1,2,3,4,5\}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are four possible domains.
(b) Answers will vary. Any set that contains all the ordered pairs' second components will do. For example, $\left\{t, t^2.t^3, t^4, t^5\right\}$, $\mathbb{P}_5(\mathbb{R})$, $\mathbb{P}_{55}(\mathbb{R})$, and $C(\mathbb{R})$ are four possible codomains.
(c) The range of a relation is the image of its domain. In this case it is $\left\{t, t^2.t^3, t^4, t^5\right\}$. You might think of it as the smallest possible codomain.

6: (a) The domain of a function is the preimage of its codomain. In this case $\{1,2,3,4,5\}$.
(b) Answers will vary. Any set that contains all the ordered pairs' second components will do. For example, $\left\{t, t^2.t^3, t^4, t^5\right\}$, $\mathbb{P}_5(\mathbb{R})$, $\mathbb{P}_{55}(\mathbb{R})$, and $C(\mathbb{R})$ are four possible codomains.
(c) The range of a function is the image of its domain. In this case it is $\left\{t, t^2.t^3, t^4, t^5\right\}$. You might think of it as the smallest possible codomain.

8: (a) The statement of the question implies that 3 is a member of the domain and we are to figure out the corresponding member of the range. This is precisely what the formula is for: $p(3) = \frac{3}{1+3^2} = \frac{3}{10}$.
(b) $p(23) = \frac{23}{1+23^2} = \frac{23}{530}$.
(c) a preimage of $\frac{2}{5}$ is any number z such that $p(z) = \frac{2}{5}$. Any solution of $\frac{2}{5} = \frac{z}{1+z^2}$ will do. The most immediate solution is $z = 2$.
(d) the preimage of $\frac{2}{5}$ is the set of all numbers z such that $p(z) = \frac{2}{5}$. Solving $\frac{2}{5} = \frac{z}{1+z^2}$:

$$\begin{aligned} 2(1+z^2) &= 5z \\ 2 - 5z + 2z^2 &= 0 \\ (2-z)(1-2z) &= 0 \end{aligned}$$

The only solutions are $z = 2, \frac{1}{2}$ so the preimage of $\frac{2}{5}$ is $\left\{\frac{1}{2}, 2\right\}$. (e) $p(\{1,2,3\}) = \left\{\frac{1}{2}, \frac{2}{5}, \frac{3}{10}\right\}$.

10: For any relation $\{(a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n)\}$ its inverse is $\{(b_1,a_1),(b_2,a_2),\ldots,(b_n,a_n)\}$. In this case, $R^{-1} = \left\{\left(\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, t^2+3\right), \left(\begin{bmatrix} 3 \\ -2 \\ 0 \end{bmatrix}, 3t^2-2t\right)\right\}$ as a subset of $\mathbb{R}^3 \times \mathbb{P}_2(\mathbb{R})$.

12b: The image of $\mathbf{a}$ means $T(\mathbf{a})$:

$$T(\mathbf{a}) = M\mathbf{a} = \begin{bmatrix} -9 & 1 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} -7 \\ 10 \end{bmatrix} = \begin{bmatrix} 73 \\ -21 \end{bmatrix}.$$

Section 4.4

3a: **[solution 1]** From the table on page 141, the elementary matrices associated with this composition are

1. reflection about the y-axis (horizontal scale by factor -1): $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$
2. vertical scale by a factor of 2: $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$
3. horizontal scale by a factor of 2: $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$

Altogether the standard matrix is the product of these elementary matrices *right to left*:

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

Note: The order of multiplication in this particular example does not matter. These particular elementary matrices commute with one another. However, not all elementary matrices commute. Generally, the order of multiplication matters.

[solution 2] The standard matrix can always be determined by finding the images of the standard basis vectors. In this case, $I_{:,1}$ maps to $\begin{bmatrix} -2 \\ 0 \end{bmatrix}$ and $I_{:,2}$ maps to $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$, verifying that indeed $\begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix}$ is the standard matrix for T.

Note: This method should be considered a double check when possible. It is not always so simple to determine the images of the standard basis vectors.

4b: Transformations with the same standard matrix are the same transformation. Determining the standard matrices for S and T (using the chart on page 141) will answer the question.

1. S: reflection about the y-axis (horizontal scaling by a factor of -1), $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$, followed by rotation $\frac{\pi}{4}$ radians clockwise ($-\frac{\pi}{4}$ radians counterclockwise) about the origin, $\begin{bmatrix} \cos\left(-\frac{\pi}{4}\right) & -\sin\left(-\frac{\pi}{4}\right) \\ \sin\left(-\frac{\pi}{4}\right) & \cos\left(-\frac{\pi}{4}\right) \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$:

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

2. T: rotation $\frac{3\pi}{4}$ radians clockwise ($-\frac{3\pi}{4}$ radians counterclockwise) about the origin, $\begin{bmatrix} \cos\left(-\frac{3\pi}{4}\right) & -\sin\left(-\frac{3\pi}{4}\right) \\ \sin\left(-\frac{3\pi}{4}\right) & \cos\left(-\frac{3\pi}{4}\right) \end{bmatrix} = \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$, followed by reflection about the x-axis (vertical scaling by a factor of -1), $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$:

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

Since S and T have the same standard matrix, they must be the same transformation (despite the differing descriptions).

Note: It is equivalent to see geometrically that the images of the standard basis vectors (the columns of the standard matrix) are the same for S and T.

5b: Given are the images of the standard basis vectors, which are the columns of the standard matrix. There is no work to do but collect the images as columns of a matrix:

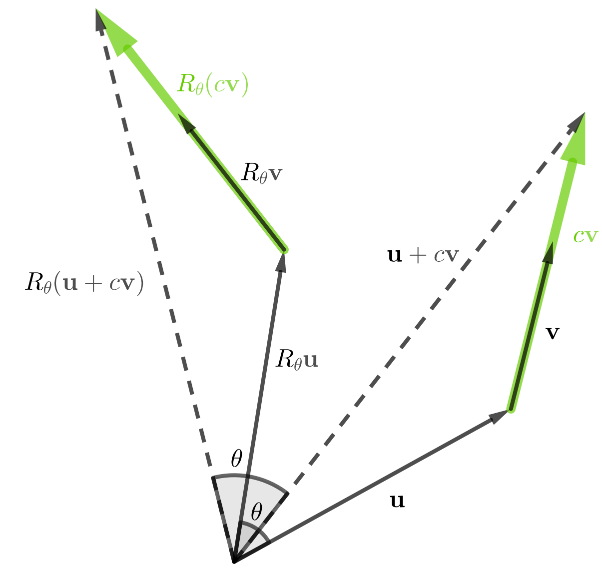
$$\begin{bmatrix} -7.5 & -5.2 \\ 1 & 10.1 \\ -13.2 & -4.3 \\ -3.7 & 4.3 \end{bmatrix}$$

7a: (**solution 1**) The columns of the standard matrix are the images of the standard basis vectors. Writing the standard basis vectors as linear combinations of the given preimages will allow computation of the images of the standard basis vectors. For example, to write $I_{:,1}$ as a linear combination of $\begin{bmatrix} 41 \\ -3 \\ -2 \end{bmatrix}, \begin{bmatrix} -6 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -5 \\ -1 \\ 1 \end{bmatrix}$ it suffices to solve

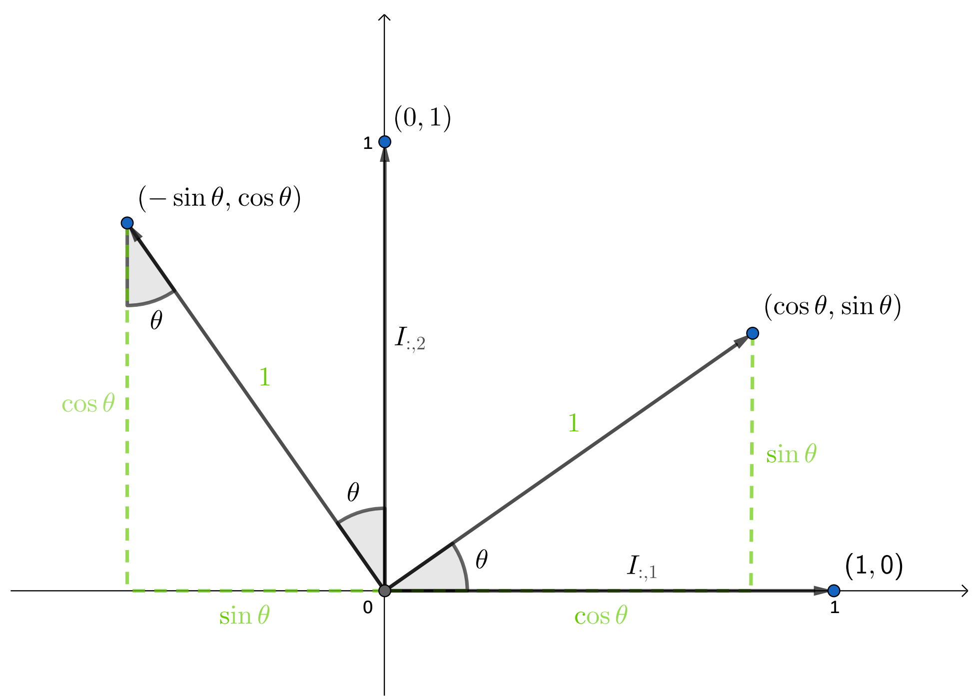
the matrix equation
$$\begin{bmatrix} 41 & -6 & -5 \\ -3 & 1 & -1 \\ -2 & 0 & 1 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$
which can be done by row reduction:
$$\begin{bmatrix} 41 & -6 & -5 & 1 \\ -3 & 1 & -1 & 0 \\ -2 & 0 & 1 & 0 \end{bmatrix} \xrightarrow{\text{reduces to}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$
Therefore,
$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1\begin{bmatrix} 41 \\ -3 \\ -2 \end{bmatrix} + 5\begin{bmatrix} -6 \\ 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} -5 \\ -1 \\ 1 \end{bmatrix}$$
and using the fact that T is a linear transformation, it then follows that
$$\begin{aligned} T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) &= 1T\left(\begin{bmatrix} 41 \\ -3 \\ -2 \end{bmatrix}\right) + 5T\left(\begin{bmatrix} -6 \\ 1 \\ 0 \end{bmatrix}\right) + 2T\left(\begin{bmatrix} -5 \\ -1 \\ 1 \end{bmatrix}\right) \\ &= 1\begin{bmatrix} 7 \\ -4 \end{bmatrix} + 5\begin{bmatrix} -13 \\ -3 \end{bmatrix} + 2\begin{bmatrix} -9 \\ -13 \end{bmatrix} = \begin{bmatrix} -76 \\ -45 \end{bmatrix}. \end{aligned}$$
The first column of the standard matrix is hence $\begin{bmatrix} -76 \\ -45 \end{bmatrix}$. Repeating the process for the second and third standard basis vectors gives the second and third columns of the standard matrix. The whole computation is neatly done in this SageMathCell 127 . The standard matrix is
$$A = \begin{bmatrix} -76 & -469 & -858 \\ -45 & -273 & -511 \end{bmatrix}$$
Double checking, we can multiply A times each given preimage to make sure it matches with the given image. For example,
$$A\begin{bmatrix} 41 \\ -3 \\ -2 \end{bmatrix} = \begin{bmatrix} -76 & -469 & -858 \\ -45 & -273 & -511 \end{bmatrix}\begin{bmatrix} 41 \\ -3 \\ -2 \end{bmatrix} = \begin{bmatrix} 7 \\ -4 \end{bmatrix}$$
as required.

(**solution 2**) We can also use the given facts more directly, and the definition of the standard matrix less directly. Given that $T\left(\begin{bmatrix} 41 \\ -3 \\ -2 \end{bmatrix}\right) = \begin{bmatrix} 7 \\ -4 \end{bmatrix}$, $T\left(\begin{bmatrix} -6 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -13 \\ -3 \end{bmatrix}$, and $T\left(\begin{bmatrix} -5 \\ -1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -9 \\ -13 \end{bmatrix}$, this may be rephrased in terms of the standard matrix A as $A \cdot \begin{bmatrix} 41 \\ -3 \\ -2 \end{bmatrix} = \begin{bmatrix} 7 \\ -4 \end{bmatrix}$, $A \cdot \begin{bmatrix} -6 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -13 \\ -3 \end{bmatrix}$, and $A \cdot \begin{bmatrix} -5 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -9 \\ -13 \end{bmatrix}$. Combining this information into a single equation,
$$A \cdot \begin{bmatrix} 41 & -6 & -5 \\ -3 & 1 & -1 \\ -2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 7 & -13 & -9 \\ -4 & -3 & -13 \end{bmatrix}.$$
Therefore $A = \begin{bmatrix} 7 & -13 & -9 \\ -4 & -3 & -13 \end{bmatrix}\begin{bmatrix} 41 & -6 & -5 \\ -3 & 1 & -1 \\ -2 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} -76 & -469 & -858 \\ -45 & -273 & -511 \end{bmatrix}$.

12: (a) As canbe seen by the following diagram showing a generic rotation, $R_\theta(\mathbf{u} + c\mathbf{v}) = R_\theta(\mathbf{u}) + cR_\theta(\mathbf{v})$.



(b) $R_\theta(I_{:,1}) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ and $R_\theta(I_{:,2}) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$ as seen below.

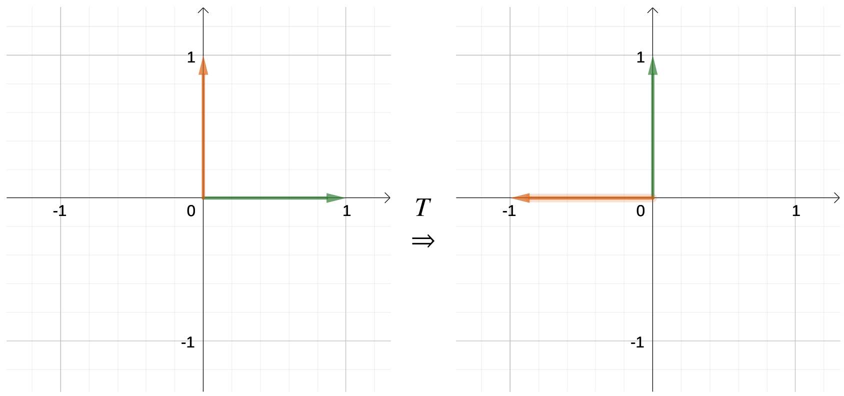


(c) $R_\theta\mathbf{u} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\mathbf{u}$

(d) Letting the arbitrary vector $\begin{bmatrix} x \\ y \end{bmatrix}$ have magnitude r and angle of incidence with the positive x-axis α:

$$\begin{aligned}
\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} r\cos\alpha \\ r\sin\alpha \end{bmatrix} = \begin{bmatrix} r\cos\theta\cos\alpha - r\sin\theta\sin\alpha \\ r\sin\theta\cos\alpha + r\cos\theta\sin\alpha \end{bmatrix} \\
&= \begin{bmatrix} r\,(\cos\theta\cos\alpha - \sin\theta\sin\alpha) \\ r\,(\sin\theta\cos\alpha + \cos\theta\sin\alpha) \end{bmatrix} = \begin{bmatrix} r\cos(\alpha+\theta) \\ r\sin(\alpha+\theta) \end{bmatrix} = R_\theta\begin{bmatrix} x \\ y \end{bmatrix}
\end{aligned}$$

13a: (i) Perhaps the simplest method for obtaining the standard matrix is to graph the action of the transformation on the standard basis vectors. Their images form the columns on the standard matrix. In this case, the following diagram demonstrates this action.



Basis vector $I_{:,1}$ maps to $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and basis vector $I_{:,2}$ maps to $\begin{bmatrix} -1 \\ 0 \end{bmatrix}$ so the standard matrix is $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

(ii) The image of $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ means $T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$, which is calculated ass

$$T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Section 4.5

1: q, h, and g are one-to-one.

q: The equation $q(a) = b$ can be solved as follows, demonstrating that it has at most one solution:

$$e^a = b \Rightarrow a = \ln(b) \text{ if } b \geq 0.$$

h: The equation $h(a) = b$ can be solved as follows, demonstrating that it has at most one solution:

$$\sqrt{a} = b \Rightarrow a = b^2 \text{ if } b \geq 0$$

g: The equation $g(a) = b$ can be solved as follows, demonstrating that it has at most one solution:

$$3x - 9 = b \Rightarrow 3x = b + 9 \Rightarrow x = \frac{b+9}{3}$$

To show that a function is not one-to-one, it is enough to provide a counterexample.
f is not one-to-one since 0 and 2π are both in the domain and $f(0) = f(2\pi) = 0$, demonstrating that the equation $f(a) = 0$ has more than one solution.
p is not one-to-one since -5 and 5 are both in the domain and $p(-5) = p(5) = 25$, demonstrating that the equation $f(a) = 25$ has more than one solution.

2: f and h are onto.

f: The equation $f(a) = b$, where b is any member of the codomain $\mathbb{R}$, is satisfied as follows, demonstrating that it has at least one solution. $a = -\sqrt[3]{\frac{b}{2}}$ is in the domain of f and

$$a = -\sqrt[3]{\frac{b}{2}} \Rightarrow a^3 = -\frac{b}{2} \Rightarrow -2a^3 = b.$$

h: The equation $h(a) = b$, where b is any member of the codomain $[0, \infty)$, can be solved as follows, demonstrating that it has at least one solution. $a = b^2$ is in the domain of h and

$$a = b^2 \Rightarrow \sqrt{a} = b.$$

To show that a function is not onto, it is enough to provide a counterexample.
q is not onto since 0 is in the codomain, but the equation $q(a) = 0$, which is to say $e^a + 1 = 0$, has no solution in the domain, $\mathbb{R}$.
p is not onto since -3 is in the codomain, but the equation $p(a) = -3$, which is to say $a^2 = -3$, has no solution in the domain, $[0, \infty)$.
g is not onto since 3 is in the codomain, but the equation $g(a) = 3$, which is to say $\cos(a) = 3$, has no solution in the domain, $\left[0, \frac{\pi}{2}\right]$.

6: Only (a). First, there are multiple elements of the domain with the same image. For example,

$$T\left(\langle 1, 2, 3, 4, 1, 2, 3, 4, \ldots\rangle\right) = 1, 2, 3, 4$$
$$\text{and}$$
$$T\left(\langle 1, 2, 3, 4, 5, 6, 7, 8, \ldots\rangle\right) = 1, 2, 3, 4$$

(and therefore the equation $T(a) = 1, 2, 3, 4$ has more than one solution, violating the definition of one-to-one).
Second, for example,

$$T\left(\langle s_1, s_2, s_3, s_4, 1, 1, 1, 1, 1, \ldots\rangle\right) = s_1, s_2, s_3, s_4$$

for any vector s_1, s_2, s_3, s_4 (and therefore the equation $T(a) = b$ has at least one solution for every element of the codomain and T is onto).
Third,

$$\begin{aligned}
T\left(\langle s_1, s_2, s_3, s_4, s_5, \ldots\rangle + c\,\langle r_1, r_2, r_3, r_4, r_5, \ldots\rangle\right)&\\
&= T\left(\langle s_1 + cr_1, s_2 + cr_2, s_3 + cr_3, s_4 + cr_4, s_5 + cr_5, \ldots\rangle\right)\\
&= s_1 + cr_1, s_2 + cr_2, s_3 + cr_3, s_4 + cr_4\\
&= s_1, s_2, s_3, s_4 + cr_1, cr_2, cr_3, cr_4\\
&= s_1, s_2, s_3, s_4 + c\left(r_1, r_2, r_3, r_4\right)
\end{aligned}$$

so T is linear.

9: (a) Yes. $T(f_1 + cf_2) = e^x(f_1 + cf_2) = e^x f_1 + ce^x f_2 = T(f_1) + cT(f_2)$. (b) Yes. Let b be an element of the codomain and suppose $T(f_1) = b$ and $T(f_2) = b$, which is to say the equation $T(a) = b$ has more than one solution. Then $T(f_1) = T(f_2)$ so $e^x f_1(x) = e^x f_2(x)$. Dividing both sides of the equation by e^x, we find that $f_1(x) = f_2(x)$. Therefore, in fact the "two solutions" must actually be one and the same (and we have shown that the equation $T(a) = b$ has at most one solution).
(c) Yes. Let b be an element of the codomain (a function) and set $f(x) = \frac{b(x)}{e^x}$. Then $T(f) = e^x \frac{b(x)}{e^x} = b(x)$ so the equation $T(a) = b$ has at least one solution (f).
(d) Yes. T is one-to-one, onto, and linear.

10c: The definitions of one-to-one and onto depend on the possible numbers of solutions of the equation $T(a) = b$, which in this case is

$$\begin{bmatrix} 14 & -1 \\ -15 & 15 \\ 12 & 9 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \tag{7.5.20}$$

(i) By theorem 5, equation (7.5.20) has at most one solution for each b precisely if the columns of the coefficient matrix are linearly independent. By inspection, we can see that they are (they are not multiples of one another), so $T(a) = b$ has at most one solution for each b in the codomain and T is one-to-one.
(ii) By theorem 6, equation (7.5.20) has at least one solution for each b precisely if the coefficient matrix has a pivot position in every row. This is, however impossible since there are 3 rows and at most 2 pivot positions, one for each column. Therefore $T(a) = b$ has no solution for some elements b of the codomain and T is not onto.
(iii) No. T is not onto and therefore is not an isomorphism.

Section 4.6

1a: The five properties defining an inner product must be verified. Direct computation will suffice. Each string of inequalities hinges on the algebra of real numbers/variables.

1. $\langle \mathbf{u}, \mathbf{u} \rangle = \left\langle \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \right\rangle = 2u_1u_1 + 3u_2u_2 = 2u_1^2 + 3u_2^2 \geq 0$
2. If $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ then $2u_1^2 + 3u_2^2 = 0$ so it must be that $u_1^2 = u_2^2 = 0$, which means $u_1 = u_2 = 0$ and therefore $\mathbf{u} = \mathbf{0}$. On the other hand, if $\mathbf{u} = \mathbf{0}$ then $u_1 = u_2 = 0$ so $\langle \mathbf{u}, \mathbf{u} \rangle = 2u_1^2 + 3u_2^2 = 0$.
3. $\langle \mathbf{u}, \mathbf{v} \rangle = \left\langle \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\rangle = 2u_1v_1 + 3u_2v_2 = 2v_1u_1 + 3v_2u_2 = \left\langle \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \right\rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
4. $\langle \mathbf{u} + \mathbf{w}, \mathbf{v} \rangle = \left\langle \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\rangle = \left\langle \begin{bmatrix} u_1 + w_1 \\ u_2 + w_2 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\rangle = 2(u_1 + w_1)v_1 + 3(u_2 + w_2)v_2$
   $= (2u_1v_1 + 3u_2v_2) + (2w_1v_1 + 3w_2v_2) = \left\langle \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\rangle + \left\langle \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle$
5.
$$\begin{aligned} \langle c\mathbf{u}, \mathbf{v} \rangle &= \left\langle c \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\rangle = \left\langle \begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\rangle = 2cu_1v_1 + 3cu_2v_2 = c(2u_1v_1 + 3u_2v_2) \\ &= c \left\langle \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\rangle = c \langle \mathbf{u}, \mathbf{v} \rangle \end{aligned}$$

10a: By definition, two vectors are orthogonal if their inner product is zero. Calculating,

$$\langle \mathbf{u}, \mathbf{v} \rangle = \left\langle \begin{bmatrix} 3 \\ 10 \end{bmatrix}, \begin{bmatrix} 10 \\ -5 \end{bmatrix} \right\rangle = \begin{bmatrix} 3 & 10 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 10 \\ -5 \end{bmatrix} = \begin{bmatrix} 3 & 10 \end{bmatrix} \begin{bmatrix} 50 \\ -15 \end{bmatrix} = 150 - 150 = 0$$

so $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

11c: By definition, two vectors are orthogonal if their inner product is zero. Calculating,

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} \rangle &= \mathbf{u}(0)\mathbf{v}(0) + \mathbf{u}(1)\mathbf{v}(1) + \mathbf{u}(2)\mathbf{v}(2) + \mathbf{u}(3)\mathbf{v}(3) \\ &= (3)(0) + (0)(-1) + (-1)(0) + (0)(3) = 0 \end{aligned}$$

so $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

12a: By definition, two vectors are orthogonal if their inner product is zero. Calculating,

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} \rangle &= \frac{1}{\pi} \int_0^{2\pi} \mathbf{u}(x)\mathbf{v}(x)\,dx = \frac{1}{\pi} \int_0^{2\pi} (\sin x)(\cos x)\,dx = -\frac{1}{2\pi} \left[\cos^2 x\right]_0^{2\pi} \\ &= -\frac{1}{2\pi} \left[\cos^2(2\pi) - \cos^2(0)\right] = -\frac{1}{2\pi} [1 - 1] = 0 \end{aligned}$$

so $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

13a: By definition, two vectors are orthogonal if their inner product is zero. Calculating,
$\mathbf{u}\mathbf{v}^T = \begin{bmatrix} 2 & 7 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} -2 & -6 \\ 8 & -4 \end{bmatrix} = \begin{bmatrix} 52 & -40 \\ 28 & 0 \end{bmatrix}$ so

$$\langle \mathbf{u}, \mathbf{v} \rangle = \left(\mathbf{u}\mathbf{v}^T\right)_{1,1} + \left(\mathbf{u}\mathbf{v}^T\right)_{2,2} = 52 + 0 = 52 \neq 0$$

so $\mathbf{u}$ and $\mathbf{v}$ are not orthogonal.

19a: Given any vector $\mathbf{v}$ in a vector space, $\mathbf{v} - \mathbf{v} = \mathbf{0}$ so

$$\begin{aligned}
\langle \mathbf{0}, \mathbf{v}\rangle &= \langle \mathbf{v} - \mathbf{v}, \mathbf{v}\rangle && \text{by substituion}\\
&= \langle \mathbf{v}, \mathbf{v}\rangle + \langle -\mathbf{v}, \mathbf{v}\rangle && \text{by inner product property 4}\\
&= \langle \mathbf{v}, \mathbf{v}\rangle + \langle (-1)\mathbf{v}, \mathbf{v}\rangle && \text{by exercise 22 of section 4.1}\\
&= \langle \mathbf{v}, \mathbf{v}\rangle + (-1)\langle \mathbf{v}, \mathbf{v}\rangle && \text{by inner product property 5}\\
&= \langle \mathbf{v}, \mathbf{v}\rangle - \langle \mathbf{v}, \mathbf{v}\rangle = 0 && \text{by algebra of real numbers}
\end{aligned}$$

Hence $\langle \mathbf{0}, \mathbf{v}\rangle = 0$ and by inner product property 3, $\langle \mathbf{v}, \mathbf{0}\rangle = \langle \mathbf{0}, \mathbf{v}\rangle$ (which of course is zero).

19c: Given any vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in a vector space,

$$\begin{aligned}
\langle \mathbf{u}, \mathbf{v} + \mathbf{w}\rangle &= \langle \mathbf{v} + \mathbf{w}, \mathbf{u}\rangle && \text{by inner product property 3}\\
&= \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{w}, \mathbf{u}\rangle && \text{by inner product property 4}\\
&= \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{u}, \mathbf{w}\rangle && \text{by inner product property 3}
\end{aligned}$$

Section 5.1

1d: The question is equivalent to asking whether $\mathbf{b}$ can be written as a linear combination of the columns of M. In turn, this is equivalent to whether $M\mathbf{v} = \mathbf{b}$ has a solution. Attempting to solve $M\mathbf{v} = \mathbf{b}$ by row reduction of the augmented matrix:

$$\begin{bmatrix} 27 & 33 & 6 \\ 9 & 11 & 2 \end{bmatrix} \xrightarrow{\frac{1}{3}A_{1,:}\to A_{1,:}} \begin{bmatrix} 9 & 11 & 2 \\ 9 & 11 & 2 \end{bmatrix} \xrightarrow{-A_{1,:}+A_{2,:}\to A_{2,:}} \begin{bmatrix} 9 & 11 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

shows that there are infinitely many solutions. Hence $\mathbf{b}$ is in the column space of M.

2d: From the echelon form $\begin{bmatrix} 9 & 11 \\ 0 & 0 \end{bmatrix}$ of M (see solution above) the only pivot position is in the first column so the set containing just the first column of M, $\left\{\begin{bmatrix} 27 \\ 9 \end{bmatrix}\right\}$ is a basis for the column space.

3d: From the echelon form $\begin{bmatrix} 9 & 11 \\ 0 & 0 \end{bmatrix}$ of M solutions of the homogeneous equation $M\mathbf{v} = \mathbf{0}$ must satisfy $9v_1 + 11v_2 = 0$ or $v_1 = -\frac{11}{9}v_2$. In parametric vector form, the solutions of $M\mathbf{v} = \mathbf{0}$ are $\mathbf{v} = r\begin{bmatrix} 11 \\ -9 \end{bmatrix}$ and therefore $\left\{\begin{bmatrix} 11 \\ -9 \end{bmatrix}\right\}$ is a basis for the null space.

4a: (i) Yes. R does not have a leading coefficient in the rightmost column. Hence the system $M\mathbf{v} = \mathbf{b}$ is consistent. (ii) The pivot positions of M are in the first, third, and fourth columns, so the first, third, and fourth columns of M comprise a basis for the column space of M:

$$\left\{\begin{bmatrix} -20 \\ -6 \\ 4 \end{bmatrix}, \begin{bmatrix} 6 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 153 \\ 45 \\ -36 \end{bmatrix}\right\}$$

Note: Since the columns of M are in $\mathbb{R}^3$ and there are three linearly indpendent columns, they must span all of $\mathbb{R}^3$ so any three linearly independent vectors form a basis for the column space of M. (iii) In parametric vector form, the solutions of $M\mathbf{v} = \mathbf{0}$ are $\mathbf{v} = r\begin{bmatrix} 9 \\ -2 \\ 0 \\ 0 \end{bmatrix}$ so one basis for the null space is

$$\left\{\begin{bmatrix} 9 \\ -2 \\ 0 \\ 0 \end{bmatrix}\right\}.$$

Note: Any set containing any nonzero multiple of $\begin{bmatrix} 9 \\ -2 \\ 0 \\ 0 \end{bmatrix}$ is also a basis for the null space.

4d: (i) No. R has a leading coefficient in the rightmost column. Hence the system $M\mathbf{v} = \mathbf{b}$ is inconsistent. (ii) The pivot positions of M are in the first, second, and fourth columns, so the first, second, and fourth columns of M comprise a basis for the column space of M:

$$\left\{ \begin{bmatrix} 187 \\ 0 \\ -33 \\ -154 \end{bmatrix}, \begin{bmatrix} 99 \\ -11 \\ -22 \\ -77 \end{bmatrix}, \begin{bmatrix} -12 \\ 4 \\ 4 \\ 12 \end{bmatrix} \right\}$$

Note: Any other set of three vectors with the same span would also do. (iii) In parametric vector form, the solutions of $M\mathbf{v} = \mathbf{0}$ are $\mathbf{v} = r\begin{bmatrix} 7 \\ -5 \\ 11 \\ 0 \end{bmatrix}$ so one basis for the null space is

$$\left\{ \begin{bmatrix} 7 \\ -5 \\ 11 \\ 0 \end{bmatrix} \right\}.$$

Note: Any set containing any nonzero multiple of $\begin{bmatrix} 7 \\ -5 \\ 11 \\ 0 \end{bmatrix}$ is also a basis for the null space.

6a: (i) Because the first and fourth columns comprise a basis for the column space, there must be a linear combination of these two columns that sums to $\mathbf{b}$. By inspection,

$$1\begin{bmatrix} -15 \\ 27 \\ 9 \end{bmatrix} + 1\begin{bmatrix} -7 \\ 42 \\ 7 \end{bmatrix} = \begin{bmatrix} -22 \\ 69 \\ 16 \end{bmatrix} = \mathbf{b}$$

(ii) From above, one particular solution of $M\mathbf{v} = \mathbf{b}$ is $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$. The solution of the homogeneous system $M\mathbf{v} = \mathbf{0}$ is given by the equations $v_1 = -\frac{7}{3}v_2 - \frac{8}{3}v_3 - \frac{7}{3}v_5$ and $v_4 = \frac{8}{7}v_5$. In parametric vector form,

$$\mathbf{v} = r\begin{bmatrix} -\frac{7}{3} \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} -\frac{8}{3} \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} -\frac{7}{3} \\ 0 \\ 0 \\ \frac{8}{7} \\ 1 \end{bmatrix}.$$

Hence the general solution of $M\mathbf{v} = \mathbf{b}$ is

$$\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + r\begin{bmatrix} -\frac{7}{3} \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} -\frac{8}{3} \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} -\frac{7}{3} \\ 0 \\ 0 \\ \frac{8}{7} \\ 1 \end{bmatrix}.$$

11: (a) **b** is -2 times the fourth column of M so $\mathbf{v} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2 \end{bmatrix}$ is one solution of $M\mathbf{v} = \mathbf{b}$.

(b) The general solution (the set of all solutions) of $M\mathbf{v} = \mathbf{b}$ is given by

$$\mathbf{v} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2 \end{bmatrix} + r \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

(c) The solution in part (a) corresponds to the general solution with $r = 0$. Picking any three other real numbers for the parameter in the general solution will provide three solutions different from the one in (a). For example, using $r = 1, 2, 3$:

$$\mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 3 \\ -2 \end{bmatrix}$$

are all solutions of $M\mathbf{v} = \mathbf{b}$.

15b: The eigenspace corresponding to λ is the null space of $M - \lambda I$. The dimension of the null space of $M - \lambda I$ is the number of free variables of the system $(M - \lambda I)\mathbf{v} = \mathbf{0}$, which can be found by row reducing $M - \lambda I$:

$$\begin{bmatrix} 25 & 45 & -75 \\ 5 & 9 & -15 \\ 15 & 27 & -45 \end{bmatrix} \underset{\frac{1}{3}A_{3,:} \to A_{3,:}}{\overset{\frac{1}{5}A_{1,:} \to A_{1,:}}{\longrightarrow}} \begin{bmatrix} 5 & 9 & -15 \\ 5 & 9 & -15 \\ 5 & 9 & -15 \end{bmatrix} \underset{-A_{1,:}+A_{3,:} \to A_{3,:}}{\overset{-A_{1,:}+A_{2,:} \to A_{2,:}}{\longrightarrow}} \begin{bmatrix} 5 & 9 & -15 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The dimension of the eigenspace is 2.

Section 5.2

5a: Finding coordinates for a vector **v** relative to a basis means finding a linear combination of the basis vectors that sums to the vector **v**. In this case we are to write (the vector) $\mathbf{v} = 3 - 4t + 5t^2$ as a linear combination of (the basis vectors) $3 - 4t$ and t^2. This can be done by inspection:

$$\mathbf{v} = 3 - 4t + 5t^2 = 1(3 - 4t) + 1(t^2)$$

so the coordinates of **v** with respect to **v** are 1 and 1 and as a coordinate vector $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\mathcal{B}}$.

6c: Finding coordinates for a vector **v** relative to a basis means finding a linear combination of the basis vectors that sums to the vector **v**. In this case we are to write (the vector) $\mathbf{v} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ as a linear combination of (the basis vectors) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}$, and $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$. This can be done by inspection:

$$\mathbf{v} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = 1\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 2\begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix} + 3\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

so the coordinates of **v** with respect to **v** are 1, 2, and 3, in that order, and as a coordinate vector $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_{\mathcal{B}}$.

7e: Finding coordinates for a vector **v** relative to a basis means finding a linear combination of the basis vectors that sums to the vector **v**. In this case we are to write (the vector) $\mathbf{v} = \langle 6, 3, -4 \rangle$ as a linear combination of (the basis vectors) $\langle 2, -1, 9 \rangle, \langle 6, 3, -4 \rangle$, and $\langle -8, 1, 1 \rangle$. This can be done by inspection:

$$\mathbf{v} = \langle 6, 3, -4 \rangle = 0\langle 2, -1, 9 \rangle + 1\langle 6, 3, -4 \rangle + 0\langle -8, 1, 1 \rangle$$

so the coordinates of $\mathbf{v}$ with respect to $\mathbf{v}$ are 0, 1, and 0, in that order, and as a coordinate vector $\mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}_{\mathcal{B}}$.

8: Finding coordinates for a vector $\mathbf{v}$ relative to a basis means finding a linear combination of the basis vectors that sums to the vector $\mathbf{v}$.

(a) We are to write (the vector) $\mathbf{v} = \begin{bmatrix} 7 \\ -12 \end{bmatrix}$ as a linear combination of (the basis vectors) $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ and $\begin{bmatrix} 5 \\ -9 \end{bmatrix}$. In other words, we need to solve the equation

$$\mathbf{v} = \begin{bmatrix} 7 \\ -12 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ -2 \end{bmatrix} + v_2 \begin{bmatrix} 5 \\ -9 \end{bmatrix}$$

which is equivalent to solving

$$\begin{bmatrix} 1 & 5 \\ -2 & -9 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 7 \\ -12 \end{bmatrix}.$$

The solution can be found by row reduction:

$$\begin{bmatrix} 1 & 5 & 7 \\ -2 & -9 & -12 \end{bmatrix} \overset{2M_{1,:}+M_{2,:}\to M_{2,:}}{\longrightarrow} \begin{bmatrix} 1 & 5 & 7 \\ 0 & 1 & 2 \end{bmatrix} \overset{-5M_{2,:}+M_{1,:}\to M_{1,:}}{\longrightarrow} \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 2 \end{bmatrix}.$$

Therefore $\begin{bmatrix} 7 \\ -12 \end{bmatrix} = -3 \begin{bmatrix} 1 \\ -2 \end{bmatrix} + 2 \begin{bmatrix} 5 \\ -9 \end{bmatrix}$ and as a coordinate vector $\mathbf{v} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}_{\mathcal{B}}$.

(b) We are to write (the vector) $\mathbf{v} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$ as a linear combination of (the basis vectors) $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ and $\begin{bmatrix} 5 \\ -9 \end{bmatrix}$. In other words, we need to solve the equation

$$\mathbf{v} = \begin{bmatrix} -2 \\ 3 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ -2 \end{bmatrix} + v_2 \begin{bmatrix} 5 \\ -9 \end{bmatrix}$$

which is equivalent to solving

$$\begin{bmatrix} 1 & 5 \\ -2 & -9 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}.$$

The solution can be found by row reduction:

$$\begin{bmatrix} 1 & 5 & -2 \\ -2 & -9 & 3 \end{bmatrix} \overset{2M_{1,:}+M_{2,:}\to M_{2,:}}{\longrightarrow} \begin{bmatrix} 1 & 5 & -2 \\ 0 & 1 & -1 \end{bmatrix} \overset{-5M_{2,:}+M_{1,:}\to M_{1,:}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \end{bmatrix}.$$

Therefore $\begin{bmatrix} -2 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ -2 \end{bmatrix} - 1 \begin{bmatrix} 5 \\ -9 \end{bmatrix}$ and as a coordinate vector $\mathbf{v} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}_{\mathcal{B}}$.

Note: There is no prescribed method for finding coordinate vectors. Any meethod of finding the desired linear combination will do, whether by inspection, row reduction, inverse matrices, or other tactic. This question could have been answered just as easily using matrix inversion, for example.

13: The columns of the change-of-basis matrix $[\mathcal{B}]_{\mathcal{E}}$ (the conversion matrix from coordinates with respect to $\mathcal{B}$ to coordinates with respect to $\mathcal{E}$) are the vectors of $\mathcal{B}$ written with respect to $\mathcal{E}$. These are given in the statement of the question: $[\mathcal{B}]_{\mathcal{E}} = \begin{bmatrix} 1 & 5 \\ -2 & -9 \end{bmatrix}$. $\begin{vmatrix} 1 & 5 \\ -2 & -9 \end{vmatrix} = 1(-9) - 5(-2) = 1$ so $[\mathcal{B}]_{\mathcal{E}}^{-1} = \begin{bmatrix} -9 & -5 \\ 2 & 1 \end{bmatrix}$.

(a) $[\mathcal{B}]_{\mathcal{E}}^{-1}\,\mathbf{v} = \begin{bmatrix} -9 & -5 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ -12 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$ which has the same coordinates as the answer for 8a.

(b) $[\mathcal{B}]_{\mathcal{E}}^{-1}\,\mathbf{v} = \begin{bmatrix} -9 & -5 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ which has the same coordinates as the answer for 8b.

Note: This process is equivalent to solving the equation $[\mathcal{B}]_{\mathcal{E}}\,\mathbf{x} = \mathbf{v}$ by matrix inversion, which gives $\mathbf{x} = [\mathcal{B}]_{\mathcal{E}}^{-1}\,\mathbf{v}$, emphasizing that $[\mathcal{B}]_{\mathcal{E}}^{-1} = [\mathcal{E}]_{\mathcal{B}}$ (the conversion matrix from coordinates with respect to $\mathcal{E}$ to coordinates with respect to $\mathcal{B}$).

19: [solution 1] The change-of-basis matrix $[\mathcal{B}]_C$ (the conversion matrix from coordinates with respect to $\mathcal{B}$ to coordinates with respect to C) is the matrix whose columns are the vectors of $\mathcal{B}$ written with respect to C. The task then is to find linear combinations of the vectors of C that sum to the vectors of $\mathcal{B}$:

$$\begin{bmatrix} -8 \\ 7 \end{bmatrix} = v_1 \begin{bmatrix} 3 \\ 7 \end{bmatrix} + v_2 \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

and

$$\begin{bmatrix} 5 \\ -6 \end{bmatrix} = v_1 \begin{bmatrix} 3 \\ 7 \end{bmatrix} + v_2 \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

These equations can be solved by row reduction, matrix inversion, or maybe even inspection. In any case, the solutions are $\begin{bmatrix} 23 \\ -\frac{77}{2} \end{bmatrix}$ and $\begin{bmatrix} -16 \\ \frac{53}{2} \end{bmatrix}$ respectively. Therefore $[\mathcal{B}]_C = \begin{bmatrix} 23 & -16 \\ -\frac{77}{2} & \frac{53}{2} \end{bmatrix}$.

25: (a) Finding $[\mathbf{v}]_{\mathcal{B}}$ means solving

$$\begin{bmatrix} 2 \\ -4 \end{bmatrix} = v_1 \begin{bmatrix} -8 \\ 7 \end{bmatrix} + v_2 \begin{bmatrix} 5 \\ -6 \end{bmatrix}.$$

By row reduction:

$$\begin{bmatrix} -8 & 5 & 2 \\ 7 & -6 & -4 \end{bmatrix} \to \begin{bmatrix} -1 & -1 & -2 \\ 7 & -6 & -4 \end{bmatrix} \to \begin{bmatrix} -1 & -1 & -2 \\ 0 & -13 & -18 \end{bmatrix}$$
$$\to \begin{bmatrix} 13 & 13 & 26 \\ 0 & -13 & -18 \end{bmatrix} \to \begin{bmatrix} 13 & 0 & 8 \\ 0 & -13 & -18 \end{bmatrix}$$

and therefore $\mathbf{v} = \begin{bmatrix} \frac{8}{13} \\ \frac{18}{13} \end{bmatrix}_{\mathcal{B}}$.

(b) Finding $[\mathbf{v}]_C$ means solving

$$\begin{bmatrix} 2 \\ -4 \end{bmatrix} = v_1 \begin{bmatrix} 3 \\ 7 \end{bmatrix} + v_2 \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

By row reduction:

$$\begin{bmatrix} 3 & 2 & 2 \\ 7 & 4 & -4 \end{bmatrix} \to \begin{bmatrix} 3 & 2 & 2 \\ 1 & 0 & -8 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -8 \\ 3 & 2 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -8 \\ 0 & 2 & 26 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -8 \\ 0 & 1 & 13 \end{bmatrix}$$

and therefore $\mathbf{v} = \begin{bmatrix} -8 \\ 13 \end{bmatrix}_C$.

(c) From question 19, $[\mathcal{B}]_C = \begin{bmatrix} 23 & -16 \\ -\frac{77}{2} & \frac{53}{2} \end{bmatrix}$ so

$$[\mathcal{B}]_C\,[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 23 & -16 \\ -\frac{77}{2} & \frac{53}{2} \end{bmatrix} \begin{bmatrix} \frac{8}{13} \\ \frac{18}{13} \end{bmatrix} = \begin{bmatrix} 23\left(\frac{8}{13}\right) - 16\left(\frac{18}{13}\right) \\ -\frac{77}{2}\left(\frac{8}{13}\right) + \frac{53}{2}\left(\frac{18}{13}\right) \end{bmatrix} = \begin{bmatrix} \frac{-104}{13} \\ \frac{338}{26} \end{bmatrix} = \begin{bmatrix} -8 \\ 13 \end{bmatrix}$$

which equals (has the same coordinates as) the answer in part (b).

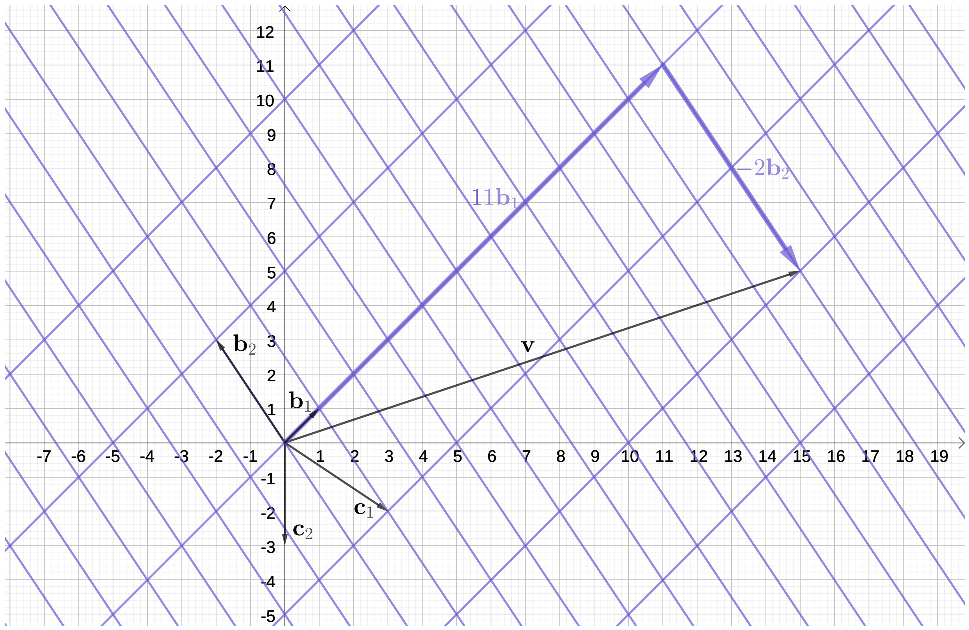
33: $[\mathcal{B}]_C$ is the conversion matrix from coordinates relative to $\mathcal{B}$ to coordinates relative to C. As a two-step process, converting from coordinates in $\mathcal{B}$ to coordinates in C can be done by converting first from coordinates in $\mathcal{B}$ to coordinates in $\mathcal{E}$ and then to coordinates in C. As a matrix equation, $[\mathcal{B}]_C = [\mathcal{E}]_C\,[\mathcal{B}]_{\mathcal{E}}$ or (because $[\mathcal{E}]_C = [C]_{\mathcal{E}}^{-1}$) $[\mathcal{B}]_C = [C]_{\mathcal{E}}^{-1}\,[\mathcal{B}]_{\mathcal{E}}$. Solving for $[C]_{\mathcal{E}}$, we find $[C]_{\mathcal{E}} = [\mathcal{B}]_{\mathcal{E}}\,[\mathcal{B}]_C^{-1}$. Hence the basis C written with respect to the standard basis appears in the columns of $[\mathcal{B}]_{\mathcal{E}}\,[\mathcal{B}]_C^{-1}$:

$$[C]_{\mathcal{E}} = \begin{bmatrix} 5 & 2 \\ 9 & -5 \end{bmatrix} \begin{bmatrix} 1 & 8 \\ -7 & 1 \end{bmatrix}^{-1} = \frac{1}{57} \begin{bmatrix} 5 & 2 \\ 9 & -5 \end{bmatrix} \begin{bmatrix} 1 & -8 \\ 7 & 1 \end{bmatrix} = \frac{1}{57} \begin{bmatrix} 19 & -38 \\ -26 & -77 \end{bmatrix}$$

and finally $C = \left\{ \frac{1}{3} - \frac{26}{57}t,\ -\frac{2}{3} - \frac{77}{57}t \right\}$.

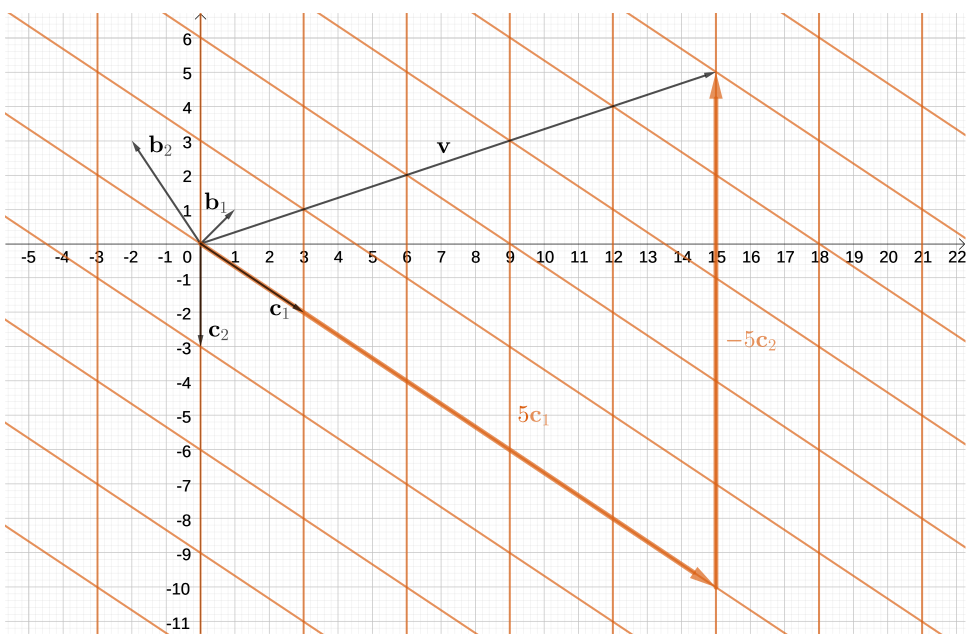
34: Finding coordinates for a vector **v** relative to a basis means finding a linear combination of the basis vectors that sums to the vector **v**. In this case, the most immediate means to a solution is graphically.

Finding $[\mathbf{v}]_{\mathcal{B}}$: the violet grid marks coordinates with respect to $\mathcal{B}$. By inspection:



$$\mathbf{v} = 11\mathbf{b}_1 - 2\mathbf{b}_2 \text{ so } \mathbf{v} = \begin{bmatrix} 11 \\ -2 \end{bmatrix}_{\mathcal{B}}.$$

Finding $[\mathbf{v}]_{C}$: the orange grid marks coordinates with respect to C. By inspection:



$$\mathbf{v} = 5\mathbf{c}_1 - 5\mathbf{c}_2 \text{ so } \mathbf{v} = \begin{bmatrix} 5 \\ -5 \end{bmatrix}_{C}.$$

Section 5.3

1c: $||\mathbf{v}|| = \sqrt{\left(\frac{15}{19}\right)^2 + \left(\frac{8}{19}\right)^2} = \frac{17}{19} \neq 1$. Since the norm of $\mathbf{v}$ is not one, $\mathbf{v}$ is not a unit vector. Scaling by the reciprocal of its norm will normalize it:

$$\frac{19}{17}\begin{bmatrix} \frac{15}{19} & \frac{8}{19} \end{bmatrix} = \begin{bmatrix} \frac{15}{17} & \frac{8}{17} \end{bmatrix}$$

so $\begin{bmatrix} \frac{15}{17} & \frac{8}{17} \end{bmatrix}$ is the normalized vector (it is a scaled version of $\mathbf{v}$ with norm 1).

2b: $\begin{bmatrix} 7 \\ 7 \end{bmatrix} \cdot \begin{bmatrix} -8 \\ 4 \end{bmatrix} = 7(-8) + 7(4) = -28 \neq 0$ so the vectors are not orthogonal and therefore the set is not orthogonal.

2g: $\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} -\frac{1}{2} \\ -5 \\ 4 \end{bmatrix} = -1 - 15 + 16 = 0$ so $\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} -\frac{1}{2} \\ -5 \\ 4 \end{bmatrix}$ are orthogonal. However, it is not enough that one pair of vectors is orthogonal. All pairs must be orthogonal for the set to be orthogonal. Checking the other two pairs, $\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 16 \\ -5 \\ -\frac{17}{4} \end{bmatrix} = 32 - 15 - 17 = 0$ and $\begin{bmatrix} -\frac{1}{2} \\ -5 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 16 \\ -5 \\ -\frac{17}{4} \end{bmatrix} = -8 + 25 - 17 = 0$. All three possible pairs of vectors are orthogonal so the set is orthogonal.

3b: $\left\langle \begin{bmatrix} 7 \\ 7 \end{bmatrix}, \begin{bmatrix} -8 \\ 4 \end{bmatrix} \right\rangle = 7(-8) + 2(7)(4) = 0$ so the vectors are orthogonal and therefore the set is orthogonal (with respect to this inner product).

5b: $\left\langle t^2 - 2t,\ 3t^2 - t - 2 \right\rangle = 0(-2) + (-1)(0) + 0(8) = 0$ so $3t^2 - t - 2$ is orthogonal to $t^2 - 2t$.

6a: $$\text{proj}_{\mathbf{u}}\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}\mathbf{u} = \frac{\begin{bmatrix} 4 \\ -3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix}}{\begin{bmatrix} 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix}} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \frac{-1}{13}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -\frac{2}{13} \\ -\frac{3}{13} \end{bmatrix}$$

11d: Coordinates of a vector $\mathbf{v}$ relative to an orthogonal basis can be computed by projecting $\mathbf{v}$ onto each vector of the basis:

$$\frac{\begin{bmatrix} 18 \\ 5 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 8 \\ 5 \end{bmatrix}}{\begin{bmatrix} 3 \\ 8 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 8 \\ 5 \end{bmatrix}} \begin{bmatrix} 3 \\ 8 \\ 5 \end{bmatrix} = \frac{89}{98}\begin{bmatrix} 3 \\ 8 \\ 5 \end{bmatrix} \qquad \frac{\begin{bmatrix} 18 \\ 5 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 25 \\ -15 \\ 9 \end{bmatrix}}{\begin{bmatrix} 25 \\ -15 \\ 9 \end{bmatrix} \cdot \begin{bmatrix} 25 \\ -15 \\ 9 \end{bmatrix}} \begin{bmatrix} 25 \\ -15 \\ 9 \end{bmatrix} = \frac{366}{931}\begin{bmatrix} 25 \\ -15 \\ 9 \end{bmatrix}$$

and

$$\frac{\begin{bmatrix} 18 \\ 5 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}}{\begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}} \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix} = \frac{69}{38}\begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$$

so $\mathbf{v} = \begin{bmatrix} \frac{89}{98} \\ \frac{366}{931} \\ \frac{69}{38} \end{bmatrix}_{\mathcal{B}}$.

12a: (**solution 1**) Following the orthogonalization procedure, the first vector of the orthogonal set is the first vector of the given set, $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. The second vector of the orthogonal set is the second vector minus its projection onto the first vector,

$$\begin{bmatrix} 5 \\ -7 \end{bmatrix} - \frac{\begin{bmatrix} 5 \\ -7 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix}}{\begin{bmatrix} 1 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 5 \\ -7 \end{bmatrix} - \frac{12}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

Therefore the result of orthogonalization is $\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right\}$.

(**solution 1**) The question did not request a specific orthogonal set. By observation, the span of the given set is $\mathbb{R}^2$ so any orthogonal set spanning $\mathbb{R}^2$ will suffice. For example, the standard basis $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ or $\left\{ \begin{bmatrix} 5 \\ 3 \end{bmatrix}, \begin{bmatrix} -6 \\ 10 \end{bmatrix} \right\}$ or any other set containing a pair of vectors whose dot product is zero.

13a: The first vector of the orthogonal set is the first vector of the given set, $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$. However, an orthonormal set is requested, so this vector needs to be scaled to a unit vector:

$$\frac{1}{\sqrt{\begin{bmatrix} 3 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix}}} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \frac{1}{\sqrt{9+16}} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix}.$$

The second vector of the orthogonal set is the second vector minus its projection onto the first, normalized, vector:

$$\begin{bmatrix} -5 \\ 10 \end{bmatrix} - \left(\begin{bmatrix} -5 \\ 10 \end{bmatrix} \cdot \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \right) \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} = \begin{bmatrix} -5 \\ 10 \end{bmatrix} - 5 \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} = \begin{bmatrix} -8 \\ 6 \end{bmatrix}.$$

Notice the calculation of the projection is slightly simplified since the denominator must be 1 as we are projecting onto a unit vector. To finish, the new vector needs to be scaled to a unit vector:

$$\frac{1}{\sqrt{\begin{bmatrix} -8 \\ 6 \end{bmatrix} \cdot \begin{bmatrix} -8 \\ 6 \end{bmatrix}}} \begin{bmatrix} -8 \\ 6 \end{bmatrix} = \frac{1}{\sqrt{64+36}} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix}.$$

Therefore the result of orthonormalization is $\left\{ \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix}, \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \right\}$.

16b: According to the solution of question 2b, the vectors are not orthogonal. Orthogonalization of a single pair of vectors amounts to subtracting from one of the vectors its projection onto the other. Letting $\mathbf{v} = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$ and $\mathbf{w} =, \begin{bmatrix} -8 \\ 4 \end{bmatrix}$,

$$\mathbf{w} - \mathrm{proj}_{\mathbf{v}} \mathbf{w} = \begin{bmatrix} -8 \\ 4 \end{bmatrix} - \frac{\begin{bmatrix} -8 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 7 \\ 7 \end{bmatrix}}{\begin{bmatrix} 7 \\ 7 \end{bmatrix} \cdot \begin{bmatrix} 7 \\ 7 \end{bmatrix}} \begin{bmatrix} 7 \\ 7 \end{bmatrix} = \begin{bmatrix} -8 \\ 4 \end{bmatrix} - \left(\frac{-56+28}{49+49} \right) \begin{bmatrix} 7 \\ 7 \end{bmatrix} = \begin{bmatrix} -6 \\ 6 \end{bmatrix}$$

Hence an orthogonal set with the same span as S is $\left\{ \begin{bmatrix} 7 \\ 7 \end{bmatrix}, \begin{bmatrix} -6 \\ 6 \end{bmatrix} \right\}$.

Normalizing each vector,

$$\frac{1}{\sqrt{\begin{bmatrix} 7 \\ 7 \end{bmatrix} \cdot \begin{bmatrix} 7 \\ 7 \end{bmatrix}}} \begin{bmatrix} 7 \\ 7 \end{bmatrix} = \frac{1}{7\sqrt{2}} \begin{bmatrix} 7 \\ 7 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

and

$$\frac{1}{\sqrt{\begin{bmatrix}-6\\6\end{bmatrix}\cdot\begin{bmatrix}-6\\6\end{bmatrix}}}\begin{bmatrix}-6\\6\end{bmatrix}=\frac{1}{6\sqrt{2}}\begin{bmatrix}-6\\6\end{bmatrix}=\begin{bmatrix}-\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{bmatrix}$$

and an orthonormal set with the same span is S is $\left\{\begin{bmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{bmatrix},\begin{bmatrix}-\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{bmatrix}\right\}$.

17b: According to the solution of question 7.5b, the vectors are orthogonal so the set S itself is an orthogonal set with the same span! Normalizing each vector,

$$\frac{1}{\sqrt{\left\langle\begin{bmatrix}7\\7\end{bmatrix},\begin{bmatrix}7\\7\end{bmatrix}\right\rangle}}\begin{bmatrix}7\\7\end{bmatrix}=\frac{1}{7\sqrt{3}}\begin{bmatrix}7\\7\end{bmatrix}=\begin{bmatrix}\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{bmatrix}$$

and

$$\frac{1}{\sqrt{\left\langle\begin{bmatrix}-6\\6\end{bmatrix},\begin{bmatrix}-6\\6\end{bmatrix}\right\rangle}}\begin{bmatrix}-6\\6\end{bmatrix}=\frac{1}{6\sqrt{3}}\begin{bmatrix}-6\\6\end{bmatrix}=\begin{bmatrix}-\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{bmatrix}$$

and an orthonormal set with the same span is S is $\left\{\begin{bmatrix}\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{bmatrix},\begin{bmatrix}-\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{bmatrix}\right\}$.

Section 5.4

1a: Solutions may vary. Any matrix P with n linearly independent eigenvectors of M will diagonalize an $n\times n$ matrix M. In this case, the simplest answer is to list the eigenvectors in the order given as the columns of P. That is, $P=\begin{bmatrix}2&1\\5&-2\end{bmatrix}$. Now

$$P^{-1}MP=\begin{bmatrix}2&1\\5&-2\end{bmatrix}^{-1}\begin{bmatrix}5&4\\20&7\end{bmatrix}\begin{bmatrix}2&1\\5&-2\end{bmatrix}=-\frac{1}{9}\begin{bmatrix}-2&-1\\-5&2\end{bmatrix}\begin{bmatrix}5&4\\20&7\end{bmatrix}\begin{bmatrix}2&1\\5&-2\end{bmatrix}$$

$$=-\frac{1}{9}\begin{bmatrix}-30&-15\\15&-6\end{bmatrix}\begin{bmatrix}2&1\\5&-2\end{bmatrix}=-\frac{1}{9}\begin{bmatrix}-135&0\\0&27\end{bmatrix}=\begin{bmatrix}15&0\\0&-3\end{bmatrix}$$

1g: Solutions may vary. Any matrix P with n linearly independent eigenvectors of M will diagonalize an $n\times n$ matrix M. In this case, the simplest answer is to list the eigenvectors in the order given as the columns of P. That is, $P=\begin{bmatrix}2&5&-5\\2&2&1\\1&0&2\end{bmatrix}$. Now

$$P^{-1}MP=\begin{bmatrix}2&5&-5\\2&2&1\\1&0&2\end{bmatrix}^{-1}\begin{bmatrix}11&-20&30\\8&-17&30\\4&-10&18\end{bmatrix}\begin{bmatrix}2&5&-5\\2&2&1\\1&0&2\end{bmatrix}$$

$$=\frac{1}{3}\begin{bmatrix}4&-10&15\\-3&9&-12\\-2&5&-6\end{bmatrix}\begin{bmatrix}11&-20&30\\8&-17&30\\4&-10&18\end{bmatrix}\begin{bmatrix}2&5&-5\\2&2&1\\1&0&2\end{bmatrix}$$

$$=\frac{1}{3}\begin{bmatrix}4&-10&15\\-3&9&-12\\-2&5&-6\end{bmatrix}\begin{bmatrix}12&15&-15\\12&6&3\\6&0&6\end{bmatrix}=\frac{1}{3}\begin{bmatrix}18&0&0\\0&9&0\\0&0&9\end{bmatrix}=\begin{bmatrix}6&0&0\\0&3&0\\0&0&3\end{bmatrix}$$

2a: The entries on the diagonal of the (diagonalized) matrix $P^{-1}MP$ are the eigenvalues of M. In this case, -3 and 15.

2g: The entries on the diagonal of the (diagonalized) matrix $P^{-1}MP$ are the eigenvalues of M. In this case, 3 and 6.

3d: Yes. An $n \times n$ matrix is diagonalizable if and only if it has a linearly independent set of n eigenvectors. In this question we have a 2×2 matrix so we need 2 linearly independent eigenvectors. Eigenvectors corresponding to different eigenvalues are necessarily linearly independent. Can you see why? Hence, without calculating them, we know the matrix has two linearly independent eigenvectors (corresponding to the two eigenvalues).

3e: An $n\times n$ matrix is diagonalizable if and only if it has a linearly independent set of n eigenvectors. In this question we have a 3×3 matrix so we need 3 linearly independent eigenvectors. Eigenvectors corresponding to different eigenvalues are necessarily linearly independent. Can you see why? Hence, without calculating them, we know the matrix has two linearly independent eigenvectors (corresponding to the two eigenvalues). We must determine whether there is a third. Reducing $M - 2I$:

$$\begin{bmatrix} 16 & -8 & 14 \\ -8 & 4 & -7 \\ -32 & 16 & -28 \end{bmatrix} \to \begin{bmatrix} -8 & 4 & -7 \\ 16 & -8 & 14 \\ -32 & 16 & -28 \end{bmatrix} \to \begin{bmatrix} -8 & 4 & -7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

we see that the equation $(M - 2I)\mathbf{v} = \mathbf{0}$ has two free variables (so the dimension of the eigenspace corresponding with 2 is two) and there are two linearly independent eigenvectors corresponding with the eigenvalue 2. Combined with an eigenvector corresponding with eigenvalue -6, we have a total of three linearly independent eigenvectors and M is diagonalizable. Yes.

4b: Similar matrices have the same determinant, so we can solve for k via

$$\begin{vmatrix} 5 & -2 \\ -4 & 3 \end{vmatrix} = \begin{vmatrix} -3 & -2 \\ 20 & k \end{vmatrix},$$

which simplifies to $15 - 8 = -3k + 40$, the solution of which is $k = 11$.

5b: Answers will vary. We are to determine matrix $P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that $PA = BP$. We found that $k = 11$ previously, so we need to solve the equation

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 5 & -2 \\ -4 & 3 \end{bmatrix} = \begin{bmatrix} -3 & -2 \\ 20 & 11 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

or

$$\begin{bmatrix} 5a - 4b & -2a + 3b \\ 5c - 4d & -2c + 3d \end{bmatrix} = \begin{bmatrix} -3a - 2c & -3b - 2d \\ 20a + 11c & 20b + 11d \end{bmatrix}$$

Matching entries gives a homogeneous linear system of four equations in the four unknowns:

$$\begin{aligned} 5a - 4b &= -3a - 2c \\ -2a + 3b &= -3b - 2d \\ 5c - 4d &= 20a + 11c \\ -2c + 3d &= 20b + 11d \end{aligned}$$

solvable by row reduction:

$$\begin{bmatrix} 8 & -4 & 2 & 0 \\ -2 & 6 & 0 & 2 \\ -20 & 0 & -6 & -4 \\ 0 & -20 & -2 & -8 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & \frac{3}{10} & \frac{1}{5} \\ 0 & 1 & \frac{1}{10} & \frac{2}{5} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

c and d are free variables, so there are many solutions. One solution is $c = 0$, $d = 5$, which makes $a = -1$ and $b = -2$. Hence $P = \begin{bmatrix} -1 & -2 \\ 0 & 5 \end{bmatrix}$ is one solution. For this choice of P, $PA = BP = \begin{bmatrix} 3 & -4 \\ -20 & 15 \end{bmatrix}$.

8a: Since similar matrices have the same determinant, eigenvalues, characteristic polynomial, and rank, we may explain their non-similarity by showing they differ on any one of these characteristics. For this particular pair, for example, they do not have the same rank. Matrix A has linearly independent columns (rank 2) while B has linearly dependent columns (rank 1). This is enough to show they are not similar. However, if we happened not to notice the differing ranks, we could also note that their determinants are not equal: $|A| = \begin{vmatrix} -6 & -1 \\ 11 & -2 \end{vmatrix} = 12 - 11 = 1$ while $|B| = \begin{vmatrix} 5 & -10 \\ -4 & 8 \end{vmatrix} = 40 - 40 = 0$. Of course we could also note that their characteristic polynomials or eigenvalues differ, but these are likely more tedious to show.

10a: Add the last two lines of the code in question 9. The result is

```
[
[a == 9*r1, b == -1/9*r1 + 1/9*r2, c == r2, d == r1]
]
```

so the solution is $P = \begin{bmatrix} 9r_1 & -\frac{1}{9}r_1 + \frac{1}{9}r_2 \\ r_2 & r_1 \end{bmatrix}$.

13a: Computing an eigenvector corresponding to $\lambda = \frac{5}{6}$:

$$\begin{bmatrix} -\frac{22}{6} & -\frac{11}{6} \\ \frac{11}{3} & \frac{11}{6} \end{bmatrix} \to \begin{bmatrix} -22 & -11 \\ 22 & 11 \end{bmatrix} \to \begin{bmatrix} -22 & -11 \\ 0 & 0 \end{bmatrix}$$

so one eigenvector is $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$. Computing an eigenvector corresponding to $\lambda = -1$:

$$\begin{bmatrix} -\frac{11}{6} & -\frac{11}{6} \\ \frac{11}{3} & \frac{11}{3} \end{bmatrix} \to \begin{bmatrix} -11 & -11 \\ 11 & 11 \end{bmatrix} \to \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

so one eigenvector is $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Letting $P = \begin{bmatrix} 1 & 1 \\ -2 & -1 \end{bmatrix}$, $P^{-1}MP = \begin{bmatrix} \frac{5}{6} & 0 \\ 0 & -1 \end{bmatrix}$ so $P^{-1}M^7P = \begin{bmatrix} \left(\frac{5}{6}\right)^7 & 0 \\ 0 & -1 \end{bmatrix}$ and therefore

$$M^7 = P\begin{bmatrix} \left(\frac{5}{6}\right)^7 & 0 \\ 0 & -1 \end{bmatrix}P^{-1} = \begin{bmatrix} 1 & 1 \\ -2 & -1 \end{bmatrix}\begin{bmatrix} \left(\frac{5}{6}\right)^7 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} -1 & -1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} -2 - \left(\frac{5}{6}\right)^7 & -1 - \left(\frac{5}{6}\right)^7 \\ 2 + 2\left(\frac{5}{6}\right)^7 & 1 + 2\left(\frac{5}{6}\right)^7 \end{bmatrix}.$$

Section 6.1

1e: Answers will vary (depending on the steps taken to reduce M). Row reducing M to an echelon form without using row swaps produces U:

$$\begin{bmatrix} 4 & 24 & 24 \\ 1 & -30 & 0 \end{bmatrix} \xrightarrow{\frac{1}{4}M_{1,:} \to M_{1,:}} \begin{bmatrix} 1 & 6 & 6 \\ 1 & -30 & 0 \end{bmatrix} \xrightarrow{-M_{1,:}+M_{2,:} \to M_{2,:}} \begin{bmatrix} 1 & 6 & 6 \\ 0 & -36 & -6 \end{bmatrix}$$

We have arrived at an echelon form, so U is determined. Applying the inverse elementary operations to the identity matrix to find L:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \xrightarrow{M_{1,:}+M_{2,:} \to M_{2,:}} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \xrightarrow{4M_{1,:} \to M_{1,:}} \begin{bmatrix} 4 & 0 \\ 1 & 1 \end{bmatrix}$$

We hence have the decomposition $M = \begin{bmatrix} 4 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 6 & 6 \\ 0 & -36 & -6 \end{bmatrix}$ (which can be double-checked by performing the multiplication).

2c: Answers will vary (depending on the steps taken to reduce M). Row reducing M to an echelon form using at least one row swap produces U:

$$\begin{bmatrix} -7 & 3 \\ 6 & -10 \\ 8 & -5 \end{bmatrix} \xrightarrow{\frac{1}{2}M_{2,:}\to M_{2,:}} \begin{bmatrix} -7 & 3 \\ 3 & -5 \\ 8 & -5 \end{bmatrix} \xrightarrow{M_{1,:}\leftrightarrow M_{2,:}} \begin{bmatrix} 3 & -5 \\ -7 & 3 \\ 8 & -5 \end{bmatrix} \xrightarrow[3M_{3,:}\to M_{3,:}]{3M_{2,:}\to M_{2,:}} \begin{bmatrix} 3 & -5 \\ -21 & 9 \\ 24 & -15 \end{bmatrix}$$

$$\xrightarrow[-8M_{1,:}+M_{3,:}\to M_{3,:}]{7M_{1,:}+M_{2,:}\to M_{2,:}} \begin{bmatrix} 3 & -5 \\ 0 & -26 \\ 0 & 25 \end{bmatrix} \xrightarrow{\frac{25}{26}M_{2,:}+M_{3,:}\to M_{3,:}} \begin{bmatrix} 3 & -5 \\ 0 & -26 \\ 0 & 0 \end{bmatrix}$$

We have arrived at an echelon form, so U is determined. Applying the inverse elementary operations to the identity matrix to find PL:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{-\frac{25}{26}M_{2,:}+M_{3,:}\to M_{3,:}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{25}{26} & 1 \end{bmatrix} \xrightarrow[8M_{1,:}+M_{3,:}\to M_{3,:}]{-7M_{1,:}+M_{2,:}\to M_{2,:}} \begin{bmatrix} 1 & 0 & 0 \\ -7 & 1 & 0 \\ 8 & -\frac{25}{26} & 1 \end{bmatrix} \xrightarrow[\frac{1}{3}M_{3,:}\to M_{3,:}]{\frac{1}{3}M_{2,:}\to M_{2,:}} \begin{bmatrix} 1 & 0 & 0 \\ -\frac{7}{3} & \frac{1}{3} & 0 \\ \frac{8}{3} & -\frac{25}{78} & \frac{1}{3} \end{bmatrix}$$

$$\xrightarrow{M_{1,:}\leftrightarrow M_{2,:}} \begin{bmatrix} -\frac{7}{3} & \frac{1}{3} & 0 \\ 1 & 0 & 0 \\ \frac{8}{3} & -\frac{25}{78} & \frac{1}{3} \end{bmatrix} \xrightarrow{2M_{2,:}\to M_{2,:}} \begin{bmatrix} -\frac{7}{3} & \frac{1}{3} & 0 \\ 2 & 0 & 0 \\ \frac{8}{3} & -\frac{25}{78} & \frac{1}{3} \end{bmatrix}$$

There is only one permutation in the row reduction, $M_{1,:} \leftrightarrow M_{2,:}$, so $P^{-1} = P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. We hence have the decomposition $M = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ -\frac{7}{3} & \frac{1}{3} & 0 \\ \frac{8}{3} & -\frac{25}{78} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 3 & -5 \\ 0 & -26 \\ 0 & 0 \end{bmatrix}$ (which can be double-checked by performing the multiplication).

3e: M is not square and therefore does not have a determinant. If M were square, the LU decomposition would give the determinant very simply. $\det(LU) = \det L \cdot \det U$. Since L and U are triangular, their determinants are simply the products of the entries on the main diagonals.

4c: M is not square and therefore does not have a determinant. If M were square, the PLU decomposition would give the determinant very simply. $\det(PLU) = \det P \cdot \det L \cdot \det U$. Since L and U are triangular, their determinants are simply the products of the entries on the main diagonals. Since P is a permutation matrix, its determinant is either 1 or -1, depending on whether it incorporates and even number or an odd number of swaps.

5e: Row reducing M to an echelon form without using row swaps or row scaling produces U:

$$\begin{bmatrix} 4 & 24 & 24 \\ 1 & -30 & 0 \end{bmatrix} \xrightarrow{-\frac{1}{4}M_{1,:}+M_{2,:}\to M_{2,:}} \begin{bmatrix} 4 & 24 & 24 \\ 0 & -36 & -6 \end{bmatrix}$$

We have arrived at an echelon form, so U is determined. Applying the inverse elementary operations to the identity matrix to find L:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \xrightarrow{\frac{1}{4}M_{1,:}+M_{2,:}\to M_{2,:}} \begin{bmatrix} 1 & 0 \\ \frac{1}{4} & 1 \end{bmatrix}$$

We hence have the decomposition $M = \begin{bmatrix} 1 & 0 \\ \frac{1}{4} & 1 \end{bmatrix} \begin{bmatrix} 4 & 24 & 24 \\ 0 & -36 & -6 \end{bmatrix}$ (which can be double-checked by performing the multiplication).

7d: Since $LU = M$, we are solving the system

$$\begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 3 & -7 & -2 \\ 0 & 7 & -1 \\ 0 & 0 & 2 \end{bmatrix} \mathbf{v} = \begin{bmatrix} 0 \\ 6 \\ 14 \end{bmatrix}$$

in two steps. First by solving $\begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \mathbf{w} = \begin{bmatrix} 0 \\ 6 \\ 14 \end{bmatrix}$ and then solving $\begin{bmatrix} 3 & -7 & -2 \\ 0 & 7 & -1 \\ 0 & 0 & 2 \end{bmatrix} \mathbf{v} = \mathbf{w}$. The vector $\mathbf{w}$ can be determined quickly by writing down the sytem represented by $\begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 14 \end{bmatrix}$ and using substitution:

$$\begin{aligned} w_1 &= 0 \\ -4w_1 + w_2 &= 6 \\ -2w_1 + 2w_2 + w_3 &= 14 \end{aligned}$$

Substituting $w_1 = 0$ into $-4w_1 + w_2 = 6$ gives $w_2 = 6$ and substituting $w_1 = 0$ and $w_2 = 6$ into $-2w_1 + 2w_2 + w_3 = 14$ gives $12 + w_3 = 14$ or $w_3 = 2$. Hence $\mathbf{w} = \begin{bmatrix} 0 & 6 & 2 \end{bmatrix}^T$ and we now need to solve $\begin{bmatrix} 3 & -7 & -2 \\ 0 & 7 & -1 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 2 \end{bmatrix}$, which can again be solved by substitution:

$$\begin{aligned} 3v_1 - 7v_2 - 2v_3 &= 0 \\ 7v_2 - v_3 &= 6 \\ 2v_3 &= 2 \end{aligned}$$

Starting with $2v_3 = 2$ we have $v_3 = 1$. Substituting $v_3 = 1$ into $7v_2 - v_3 = 6$ yields $7v_2 = 7$ or $v_2 = 1$. Substituting $v_3 = 1$ and $v_2 = 1$ into $3v_1 - 7v_2 - 2v_3 = 0$ yields $3v_1 = 9$ or $v_1 = 3$. Hence the solution of

$$\begin{bmatrix} 3 & -7 & -2 \\ -12 & 35 & 7 \\ -6 & 28 & 4 \end{bmatrix} \mathbf{v} = \begin{bmatrix} 0 \\ 6 \\ 14 \end{bmatrix} \quad \text{is} \quad \mathbf{v} = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}.$$

Section 6.2

1b: The eigenvalues of M are determined by its characteristic equation:

$$|M - \lambda I| = \begin{vmatrix} -4-\lambda & 2 \\ -3 & -9-\lambda \end{vmatrix} = (-4-\lambda)(-9-\lambda) + 6 = \lambda^2 + 13\lambda + 42 = 0$$

which can be factored as $(\lambda + 7)(\lambda + 6) = 0$ giving eigenvalues -7 and -6. The magnitudes (absolute values) of the eigenvalues are 7 and 6, so -7 is the dominant eigenvalue (the one with greatest magnitude).

2b: The eigenvalues of M are determined by its characteristic equation:

$$|M - \lambda I| = \begin{vmatrix} 19-\lambda & 12 \\ -28 & -19-\lambda \end{vmatrix} = (19-\lambda)(-19-\lambda) + 336 = \lambda^2 - 25 = 0$$

which can be factored as $(\lambda+5)(\lambda-5) = 0$ giving eigenvalues -5 and 5. The magnitudes (absolute values) of the eigenvalues are both 5, so there is no dominant eigenvalue (one with greatest magnitude). The two eigenvalues have the same magnitude.

3a: The most immediate approximation of an eigenvector is either $\mathbf{v}_5$ or $\mathbf{v}_6$ itself. The simplest way to get a handle on the associated eigenvalue is to note that $\mathbf{v}_6 = M\mathbf{v}_5$, which should be approximately $\lambda\mathbf{v}_5$.

$$\mathbf{v}_6 = \begin{bmatrix} 1894 \\ -3020 \end{bmatrix} \approx \lambda \begin{bmatrix} -467 \\ 742 \end{bmatrix} = \begin{bmatrix} -467\lambda \\ 742\lambda \end{bmatrix}$$

Matching entries, we see that $1894 \approx -467\lambda$ and $-3020 \approx 742\lambda$, and solving for λ, $\lambda \approx \frac{1894}{-467} \approx -4.056$ and $\lambda \approx \frac{-3020}{742} \approx -4.070$. Either one of these estimates will do for our approximation. There is no reason to choose one over the other save personal preference. As for an approximate eigenvector, we expect $\mathbf{v}_6$ to be a better approximation than $\mathbf{v}_5$. In summary, we choose

$$-4.065, \begin{bmatrix} 1894 \\ -3020 \end{bmatrix}$$

as the approximate eigenpair.

4a: The code computes $\mathbf{v}_1$ through $\mathbf{v}_{11}$ for us. Since the ratios $(\mathbf{v}_{10})_{1,1} : (\mathbf{v}_{10})_{2,1}$ and $(\mathbf{v}_{11})_{1,1} : (\mathbf{v}_{11})_{2,1}$ are $5196345 : 12990681$ and $171478659 : 428694651$, or approximately $1 : 2.49997$ and $1 : 2.49999$ it seems that the eigenvector is converging to some multiple of $\begin{bmatrix} 1 \\ 2.5 \end{bmatrix}$ and therefore yes, it seems as if the method will converge.

5a: The code computes $\mathbf{v}_6$ through $\mathbf{v}_{100}$ for us. Since $\mathbf{v}_{100} = \begin{bmatrix} 0.625 \\ -1 \end{bmatrix}$ and $\mathbf{v}_{52}$ through $\mathbf{v}_{99}$ are multiples of this vector (1 or -1 times), the method has settled/converged on this direction as an eigenvector. To find the associated eigenvalue, we compute

$$M\mathbf{v}_{100} = \begin{bmatrix} -12 & -5 \\ 16 & 6 \end{bmatrix} \begin{bmatrix} 0.625 \\ -1 \end{bmatrix} = \begin{bmatrix} -2.5 \\ 4 \end{bmatrix}$$

Noting that $\begin{bmatrix} -2.5 \\ 4 \end{bmatrix} = -4 \begin{bmatrix} 0.625 \\ -1 \end{bmatrix}$, the associated eigenvalue is -4 and we have the exact eigenpair

$$-4, \begin{bmatrix} 0.625 \\ -1 \end{bmatrix}.$$

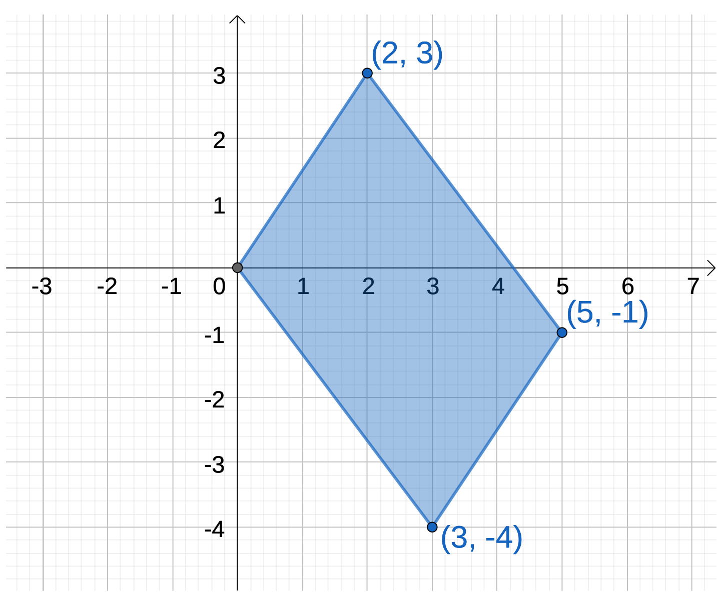
Note: This compares well to the approximation from question 3a where the result was

$$-4.065, \begin{bmatrix} 1894 \\ -3020 \end{bmatrix}.$$

Scaling the approximate eigenvector by $\frac{1}{3020}$ gives (a scaled eigenvector approximation of about) $\begin{bmatrix} 0.62715 \\ -1 \end{bmatrix}$.

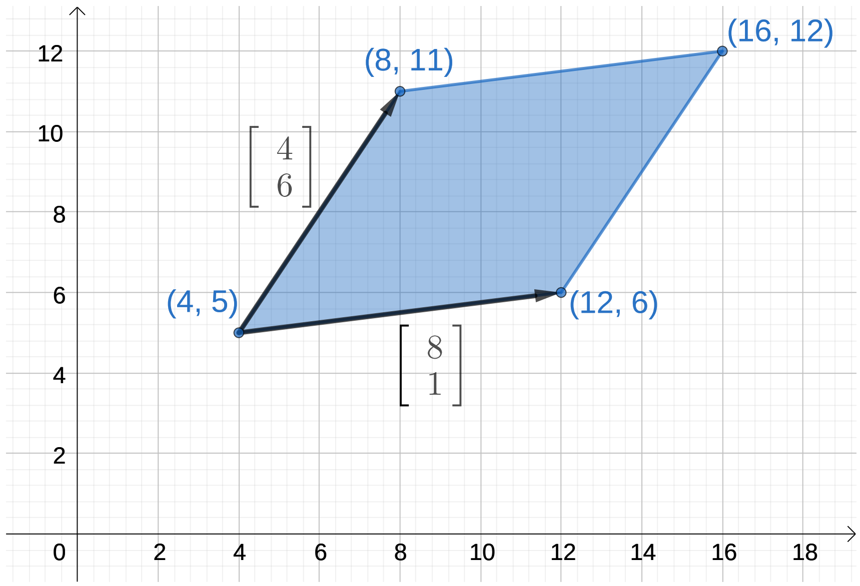
Section 6.3

1a: The parallelogram in question is sketched below.



By an analysis much like that in subsection **Areas and eigenvalues**, an argument can be made that this parallelogram is the image of the unit square (the square with vertices at $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$) under the linear transformation $T_A(\mathbf{v}) = A\mathbf{v} = \begin{bmatrix} 2 & 3 \\ 3 & -4 \end{bmatrix}\mathbf{v}$, which maps $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ to $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to $\begin{bmatrix} 3 \\ -4 \end{bmatrix}$. Letting P be the parallelogram and S the unit square, we have $T(S) = P$ and therefore $\text{area}(P) = |\det A| \cdot \text{area}(S) = 17 \cdot 1$. The area of the parallelogram is 17.

1d: The parallelogram in question is sketched below.



By an analysis much like that in subsection **Areas and eigenvalues**, an argument can be made that this parallelogram is the image of the unit square (the square with vertices at $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$) under the affine transformation $T(\mathbf{v}) = A\mathbf{v} + \mathbf{b} = \begin{bmatrix} 4 & 8 \\ 6 & 1 \end{bmatrix}\mathbf{v} + \begin{bmatrix} 4 \\ 5 \end{bmatrix}$, which maps $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ to $\begin{bmatrix} 4 \\ 5 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ to $\begin{bmatrix} 8 \\ 11 \end{bmatrix}$, and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to $\begin{bmatrix} 12 \\ 6 \end{bmatrix}$. Letting P be the parallelogram and S the unit square, we have $T(S) = P$ and therefore $\text{area}(P) = |\det A| \cdot \text{area}(S) = 44 \cdot 1$. The area of the parallelogram is 44.

3b: The area of the parallelogram determined by the columns of $\begin{bmatrix} -1 & 6 \\ 4 & -2 \end{bmatrix}$ is exactly twice the area of the given triangle. By question 2, $\left|\det \begin{bmatrix} -1 & 6 \\ 4 & -2 \end{bmatrix}\right| = 22$ is the area of the paralellogram, so the area of the triangle is 11.

6a: Because M and U are similar, they have the same determinant and eigenvalues (see Section 5.4). Because U is triangular, (i) $\det U$ is the product of the entries on its diagonal, $(-4)(5) = -20 = \det M$; and (ii) the eigenvalues of U are the entries on its diagonal: -4 and 5 are the eigenvalues of M.

7a: Adding the lines

```
print("M =")
print(M); print()
print("det M =",M.det())
print("eigenvalues are",M.eigenvalues())
```

to the SageMath code produces

```
M =
[-13/37  54/37]
```

```
[495/37  50/37]

det M = -20
eigenvalues are [5, -4]
```

Hence $\det M = -20$ and the eigenvalues of M are -5 and 4 as determined before.

11a: Following the algorithm requires finding an eigenpair, so we find an eigenvector of M corresponding to eigenvalue $\lambda = -3$:

$$M - \lambda I = \begin{bmatrix} 16 & -8 \\ 20 & -10 \end{bmatrix} \xrightarrow{-\frac{5}{4}A_{1,:}+A_{2,:}\to A_{2,:}} \begin{bmatrix} 16 & -8 \\ 0 & 0 \end{bmatrix}$$

so eigenvectors $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ satisfy $16v_1 - 8v_2 = 0$ or $v_2 = 2v_1$. Using eigenvector $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ we set $Q = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$. Since $\left(Q^{-1}MQ\right)_{\backslash 1,1}$ is a 1×1 matrix, it is already upper triangular, and we can select $R = [1]$. Hence we have $\hat{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and therefore $P = Q\hat{Q} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$.

Note that

$$P^{-1}MP = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 13 & -8 \\ 20 & -13 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 13 & -8 \\ 20 & -13 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 13 & -8 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} -3 & -8 \\ 0 & 3 \end{bmatrix},$$

which is upper triangular.

11e: Following the algorithm requires finding an eigenpair, so we find an eigenvector of M corresponding to eigenvalue $\lambda = 4$:

$$M - \lambda I = \begin{bmatrix} -38 & -50 & -24 \\ 37 & 49 & 24 \\ -26 & -35 & -18 \end{bmatrix} \underset{-\frac{4}{3}A_{3,:}+A_{1,:}\to A_{1,:}}{\xrightarrow{\frac{4}{3}A_{3,:}+A_{2,:}\to A_{2,:}}} \begin{bmatrix} -\frac{10}{3} & -\frac{10}{3} & 0 \\ \frac{7}{3} & \frac{7}{3} & 0 \\ -26 & -35 & -18 \end{bmatrix} \underset{\frac{3}{7}A_{2,:}\to A_{2,:}}{\xrightarrow{-\frac{3}{10}A_{1,:}\to A_{1,:}}}$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ -26 & -35 & -18 \end{bmatrix} \underset{35A_{2,:}+A_{3,:}\to A_{3,:}}{\xrightarrow{-A_{2,:}+A_{1,:}\to A_{1,:}}} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 9 & 0 & -18 \end{bmatrix}$$

so eigenvectors $\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ satisfy $v_2 = -v_1$ and $v_3 = \frac{1}{2}v_1$. Using eigenvector $\mathbf{v} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$ we set $Q = \begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$. Since $\left(Q^{-1}MQ\right)_{\backslash 1,1}$ is a 2×2 matrix, it must be calculated and triangularized. By row reduction, $Q^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 1 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix}$ and so

$$Q^{-1}MQ = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 1 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} -34 & -50 & -24 \\ 37 & 53 & 24 \\ -26 & -35 & -14 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -25 & -12 \\ 0 & 3 & 0 \\ 0 & -10 & -2 \end{bmatrix}$$

and $\left(Q^{-1}MQ\right)_{\backslash 1,1} = \begin{bmatrix} 3 & 0 \\ -10 & -2 \end{bmatrix}$ must now be triangularized (applying the algorithm as done for question 11a). The eigenvalues of $\left(Q^{-1}MQ\right)_{\backslash 1,1}$ are 3 and -2. Choosing to find an eigenvector for eigenvalue 3,

$\left(Q^{-1}MQ\right)_{\backslash 1,1} - 3I = \begin{bmatrix} 0 & 0 \\ -10 & -5 \end{bmatrix}$ so a corresponding eigenvector is $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$. Hence $\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$ triangularizes $\left(Q^{-1}MQ\right)_{\backslash 1,1}$ and we set $R = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$. Returning to the triangularization of M,

$$\hat{Q} = \begin{bmatrix} 1 & 0 \\ 0 & R \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \quad \text{and} \quad P = Q\hat{Q} = \begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}.$$

Multiplying, we find $P = \begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix}$.

Note that

$$\begin{aligned} P^{-1}MP &= \begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -34 & -50 & -24 \\ 37 & 53 & 24 \\ -26 & -35 & -14 \end{bmatrix}\begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 1 & 1 & 0 \\ \frac{3}{2} & 2 & 1 \end{bmatrix}\begin{bmatrix} -34 & -50 & -24 \\ 37 & 53 & 24 \\ -26 & -35 & -14 \end{bmatrix}\begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -17 & -25 & -12 \\ 3 & 3 & 0 \\ -3 & -4 & -2 \end{bmatrix}\begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -1 & -12 \\ 0 & 3 & 0 \\ 0 & 0 & -2 \end{bmatrix}, \end{aligned}$$

which is upper triangular.

Section 6.4

1b: This question is requesting the best approximation of (the point) $\mathbf{p} = \begin{bmatrix} 12 \\ 1 \end{bmatrix}$ within the subspace $W = \operatorname{span}\left\{\begin{bmatrix} 7 \\ -6 \end{bmatrix}\right\}$ (the collection of all multiples of $\mathbf{v}$). By theorem 18, the answer is the projection of $\mathbf{p}$ onto W:

$$\operatorname{proj}_W \mathbf{p} = \operatorname{proj}_{\mathbf{v}} \mathbf{p} = \frac{\mathbf{p}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v} = \frac{12(7) + 1(-6)}{7(7) + (-6)(-6)}\mathbf{v} = \frac{78}{85}\mathbf{v}.$$

2b: The distance between a point and line is the distance between the point and the nearest point on the line (the best approximation of the point on the line). The line $y = \frac{1}{5}x$ is shorthand for the set of points $\left\{(x, y) : y = \frac{1}{5}x\right\}$, or as vectors, $\left\{\begin{bmatrix} x \\ y \end{bmatrix} : y = \frac{1}{5}x\right\} = \left\{\begin{bmatrix} x \\ \frac{1}{5}x \end{bmatrix} : x \text{ in } \mathbb{R}\right\} = \left\{x\begin{bmatrix} 1 \\ \frac{1}{5} \end{bmatrix} : x \text{ in } \mathbb{R}\right\} = \operatorname{span}\left\{\begin{bmatrix} 1 \\ \frac{1}{5} \end{bmatrix}\right\} = \operatorname{span}\left\{\begin{bmatrix} 5 \\ 1 \end{bmatrix}\right\}$. Hence we seek the best approximation of (the point) $\mathbf{p} = \begin{bmatrix} -3 \\ -4 \end{bmatrix}$ within the subspace $W = \operatorname{span}\left\{\begin{bmatrix} 5 \\ 1 \end{bmatrix}\right\}$ (the line $y = \frac{1}{5}x$). By theorem 18, the nearest point to $\mathbf{p}$ is the projection of $\mathbf{p}$ onto W:

$$\operatorname{proj}_W \mathbf{p} = \frac{-3(5) - 4(1)}{5(5) + 1(1)}\mathbf{p} = -\frac{19}{26}\begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{95}{26} \\ -\frac{19}{26} \end{bmatrix}.$$

The distance between the point and line is then the distance between the point and the best approximation, the norm of their difference:

$$\left\|\begin{bmatrix} -3 \\ -4 \end{bmatrix} - \begin{bmatrix} -\frac{95}{26} \\ -\frac{19}{26} \end{bmatrix}\right\| = \left\|\begin{bmatrix} \frac{17}{26} \\ -\frac{85}{26} \end{bmatrix}\right\| = \frac{1}{26}\left\|\begin{bmatrix} 17 \\ 85 \end{bmatrix}\right\| = \frac{\sqrt{17^2 + 85^2}}{26} = \frac{\sqrt{7514}}{26} \approx 3.334$$

3e: By definition, the projection of a vector (as the given $\mathbf{v}$) onto a subspace is the sum of its projections onto an orthogonal basis (as the given $\mathcal{B}$) of the subspace.

$$\operatorname{proj}_{\operatorname{span}\mathcal{B}}\mathbf{v} = \frac{\begin{bmatrix} 9 \\ -12 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ -4 \\ 5 \end{bmatrix}}{\begin{bmatrix} -1 \\ -4 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ -4 \\ 5 \end{bmatrix}} \begin{bmatrix} -1 \\ -4 \\ 5 \end{bmatrix} + \frac{\begin{bmatrix} 9 \\ -12 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}}{\begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}} \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix} + \frac{\begin{bmatrix} 9 \\ -12 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \frac{64}{42} \begin{bmatrix} -1 \\ -4 \\ 5 \end{bmatrix} + \frac{-46}{14} \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 9 \\ -12 \\ 5 \end{bmatrix}$$

Note: This could have been determined without calculation. We have a basis with 3 elements in $\mathbb{R}^3$, so $\operatorname{span}\mathcal{B} = \mathbb{R}^3$ and therefore $\mathbf{v}$ is in $\operatorname{span}\mathcal{B}$. The vector is in the subspace in question and so does not need to be approximated. It can be represented exactly as a linear combination of any basis. See question 10.

4b: $\operatorname{span}\mathcal{B}$ equals the span of the column space of $[\mathcal{B}]$, which can be determined by row reduction:

$$\begin{bmatrix} 0 & 6 & 6 \\ -9 & -16 & -7 \\ -6 & 2 & 8 \end{bmatrix} \to \begin{bmatrix} 3 & -1 & -4 \\ -9 & -16 & -7 \\ 0 & 6 & 6 \end{bmatrix} \to \begin{bmatrix} 3 & -1 & -4 \\ 0 & -19 & -19 \\ 0 & 6 & 6 \end{bmatrix} \to \begin{bmatrix} 3 & -1 & -4 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

The first two columns have pivot positions, so the first two vectors of $\mathcal{B}$ form a basis for $\operatorname{span}\mathcal{B}$. In order to project onto $\operatorname{span}\mathcal{B}$, however, we need an orthogonal basis. Letting $\mathbf{b}_1 = -\frac{1}{3}\begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix} =$ and $\mathbf{b}_2 = \frac{1}{2}\begin{bmatrix} 6 \\ -16 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ -8 \\ 1 \end{bmatrix}$,

$$\mathbf{b}_2 - \operatorname{proj}_{\mathbf{b}_1}\mathbf{b}_2 = \begin{bmatrix} 3 \\ -8 \\ 1 \end{bmatrix} - \frac{-22}{13}\begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ -\frac{38}{13} \\ \frac{57}{13} \end{bmatrix}$$

$13\begin{bmatrix} 3 \\ -\frac{38}{13} \\ \frac{57}{13} \end{bmatrix} = \begin{bmatrix} 39 \\ -38 \\ 57 \end{bmatrix}$, so we may use $\left\{\mathbf{w}_1 = \begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 39 \\ -38 \\ 57 \end{bmatrix}\right\}$ for an orthogonal basis. Finally,

$$\operatorname{proj}_{\operatorname{span}\mathcal{B}}\mathbf{v} = \operatorname{proj}_{\mathbf{w}_1}\mathbf{v} + \operatorname{proj}_{\mathbf{w}_2}\mathbf{v} = \frac{\begin{bmatrix} 7 \\ -7 \\ 9 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix}}{\begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix}} \begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix} + \frac{\begin{bmatrix} 7 \\ -7 \\ 9 \end{bmatrix} \cdot \begin{bmatrix} 39 \\ -38 \\ 57 \end{bmatrix}}{\begin{bmatrix} 39 \\ -38 \\ 57 \end{bmatrix} \cdot \begin{bmatrix} 39 \\ -38 \\ 57 \end{bmatrix}} \begin{bmatrix} 39 \\ -38 \\ 57 \end{bmatrix}$$

$$= \frac{-3}{13}\begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix} + \frac{1052}{6214}\begin{bmatrix} 39 \\ -38 \\ 57 \end{bmatrix} = \frac{1}{239}\begin{bmatrix} 1578 \\ -1703 \\ 2196 \end{bmatrix} \approx \begin{bmatrix} 6.6025 \\ -7.1255 \\ 9.1883 \end{bmatrix}$$

5b: By definition $\mathbf{v}$ is in $W^{\perp}$ if it is orthogonal to all vectors in subspace W. Since

$$W = \left\{ a\begin{bmatrix} -7 \\ 0 \\ 0 \end{bmatrix} + b\begin{bmatrix} 8 \\ -3 \\ -4 \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\},$$

we may represent a particular but arbitrary vector $\mathbf{w}$ of W by $\mathbf{w} = a\begin{bmatrix} -7 \\ 0 \\ 0 \end{bmatrix} + b\begin{bmatrix} 8 \\ -3 \\ -4 \end{bmatrix}$ for some real numbers a and b. Then

$$\mathbf{w} \cdot \mathbf{v} = \left(a\begin{bmatrix} -7 \\ 0 \\ 0 \end{bmatrix} + b\begin{bmatrix} 8 \\ -3 \\ -4 \end{bmatrix}\right) \cdot \begin{bmatrix} 0 \\ -8 \\ 6 \end{bmatrix} = a\begin{bmatrix} -7 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -8 \\ 6 \end{bmatrix} + b\begin{bmatrix} 8 \\ -3 \\ -4 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -8 \\ 6 \end{bmatrix}$$
$$= a(-7(0) + 0(-8) + 0(6)) + b(8(0) - 3(-8) - 4(6)) = 0.$$

So $\mathbf{v}$ is orthogonal to all elements of W and is therefore in $W^\perp$.

Note: In retrospect, showing that $\mathbf{v}$ is orthogonal to all elements of the given basis of W (shown as part of the computation above) is sufficient.

6: We have determined above that $\mathbf{v}$ is in $W^\perp$, so one decomposition is $\mathbf{v} = \mathbf{0} + \begin{bmatrix} 0 \\ -8 \\ 6 \end{bmatrix}$. $\mathbf{0}$ is in W and $\begin{bmatrix} 0 \\ -8 \\ 6 \end{bmatrix}$ is in $W^\perp$ as required. (And by theorem 16, this is the only such decomposition).

7a: By row reduction, the solution of the system:

$$\begin{bmatrix} -6 & 3 & -15 & 7 \\ 14 & -7 & 35 & 1 \\ -18 & 9 & -45 & 6 \end{bmatrix} \xrightarrow{-3A_{1,:}+A_{3,:}\to A_{3,:}} \begin{bmatrix} -6 & 3 & -15 & 7 \\ 14 & -7 & 35 & 1 \\ 0 & 0 & 0 & -15 \end{bmatrix}$$

The system is inconsistent! We must find a best approximation. According to theorem 19, we need to project the constant vector onto the column space of the coefficient matrix. Continuing the row reduction begun above, but without the rightmost column:

$$\xrightarrow{\frac{7}{3}A_{1,:}+A_{2,:}\to A_{2,:}} \begin{bmatrix} -6 & 3 & -15 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The only pivot position of the coefficient matrix appears in the first column, so the first column of the coefficient matrix provides a basis for the column space. We therefore project $\begin{bmatrix} 7 \\ 1 \\ -15 \end{bmatrix}$ onto $\begin{bmatrix} -6 \\ 14 \\ -18 \end{bmatrix}$:

$$\frac{7(-6) + 1(14) - 15(-18)}{6^2 + 14^2 + 18^2}\begin{bmatrix} -6 \\ 14 \\ -8 \end{bmatrix} = \frac{121}{278}\begin{bmatrix} -6 \\ 14 \\ -8 \end{bmatrix}$$

and a best approximation of the solution is $x = \frac{121}{278}$, $y = 0$, $z = 0$.

Note: Answers may vary. Any linear combination of the columns of the coefficient matrix that sums to $\frac{121}{278}\begin{bmatrix} -6 \\ 14 \\ -8 \end{bmatrix}$ will do. There are infinitely many of them.

8b: By theorem 16, every vector of $\mathbb{R}^3$ can be written as a sum of a vector in W and a vector in $W^\perp$. Hence if $\mathcal{B}$ is a basis for W and $\mathcal{B}^\perp$ is a basis for $W^\perp$, it must be that $\mathcal{B} \cup \mathcal{B}^\perp$ spans $\mathbb{R}^3$. That $\mathcal{B} \cup \mathcal{B}^\perp$ is linearly independent (and therefore a basis is left as an exercise—see exercise 12). In the case of this question, $\mathcal{B}$ is given with one element meaning $\mathcal{B}^\perp$ must have two elements. We seek two linearly independent vectors orthogonal to $W = \operatorname{span}\left\{\begin{bmatrix} 7 \\ 11 \\ 12 \end{bmatrix}\right\}$. A quick computation will show $\begin{bmatrix} -11 \\ 7 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ -12 \\ 11 \end{bmatrix}$ are orthogonal to W (do you see where they came from?). Since $\begin{bmatrix} -11 \\ 7 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ -12 \\ 11 \end{bmatrix}$ are additionally linearly independent, they form a basis for $W^\perp$.

13c: This question is requesting the best approximation of (the point) $\mathbf{p} = x^3 - 11x^2 - 9x + 10$ within the subspace $W = \operatorname{span}\left\{12x^3 - x - 5\right\}$ (the collection of all multiples of $\mathbf{q}$). By theorem 18, the answer is the projection of $\mathbf{p}$ onto W:

$$\operatorname{proj}_W \mathbf{p} = \operatorname{proj}_{\mathbf{q}} \mathbf{p} = \frac{\langle \mathbf{p}, \mathbf{q} \rangle}{\langle \mathbf{q}, \mathbf{q} \rangle} \mathbf{q} = \frac{-4132}{8238} \mathbf{q} = -\frac{2066}{1083} \mathbf{q}.$$

14c: According to the solution above, the best approximation is $-\frac{2066}{1083}\mathbf{q}$. Distance is measured as the norm of the difference, which is $\left\| -\frac{2066}{1083}\mathbf{q} - \mathbf{p} \right\|$:

$$\left\| -\frac{2066}{1083}\mathbf{q} - \mathbf{p} \right\| = \frac{1}{1083} \left\| 2066\mathbf{q} + 1083\mathbf{p} \right\| = \frac{1}{1083} \left\| 25875x^3 - 11913x^2 - 11813x + 500 \right\|$$

$$= \frac{1}{1083} \sqrt{2,5513,748} = \frac{2\sqrt{6378437}}{1083} \approx 4.664$$

Section 7.1

3a: The coefficient matrix and constant vector for the linear regression problem are

$$M = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_8 & x_8^2 & x_8^3 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 141.1 \\ -35.51 \\ \vdots \\ 1783 \end{bmatrix}.$$

According to the code at SageMathCell 128 M is (with whitespace changes only)

```
[   1     389/1000      151321/1000000         58863869/1000000000]
[   1     851/1000      724201/1000000        616295051/1000000000]
[   1    2467/1000     6086089/1000000      15014381563/1000000000]
[   1    4113/1000    16916769/1000000      69578670897/1000000000]
[   1      181/40        32761/1600                 5929741/64000]
[   1    6639/1000    44076321/1000000     292622695119/1000000000]
[   1    8873/1000    78730129/1000000     698572434617/1000000000]
[   1      281/25        78961/625                 22188041/15625]
```

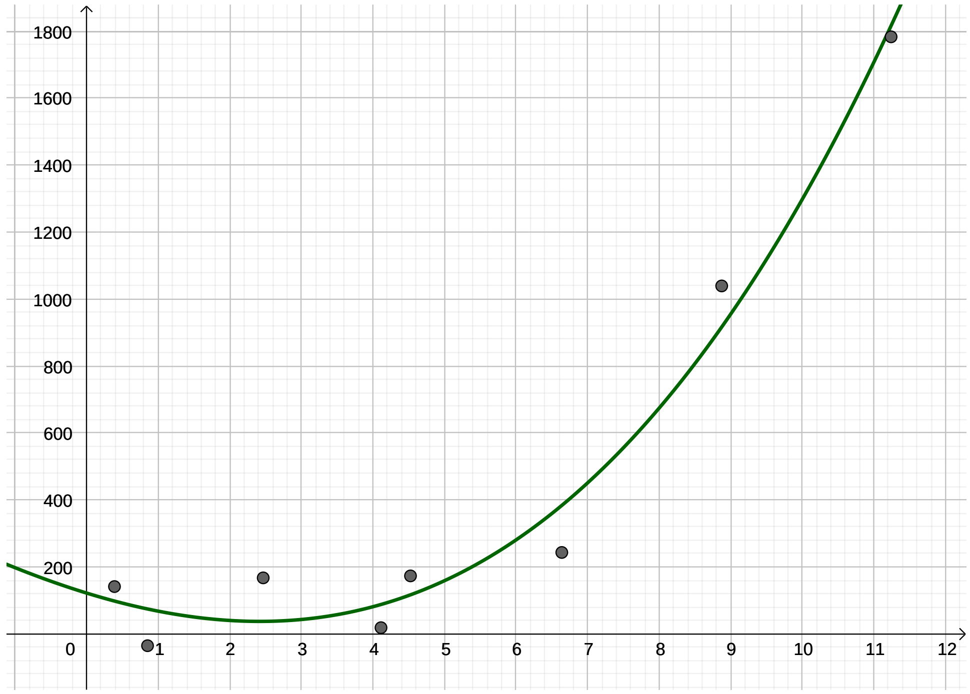
The code proceeds to calculate the normal equations $M^T M \hat{\mathbf{v}} = M^T \mathbf{b}$ and solve for $\hat{\mathbf{v}}$, the regression coefficients:

$$\begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{bmatrix} = \begin{bmatrix} 121.697796 & -65.6594829 & 10.8873569 & 0.744289915 \end{bmatrix}^T$$

Hence the best fit model is

$$f(x) = 0.744289915x^3 + 10.8873569x^2 - 65.6594829x + 121.697796.$$

A plot of $f(x)$ superimposed on a scatterplot of the data is shown below, demonstrating geometrically the goodness of fit.



Note: Unlike with the method of projection, it is not critical that the data are entered exactly nor that the computation is done using exact arithmetic. Replacing the first two lines of the code by

```
x = vector([.389,.851,2.467,4.113,4.525,6.639,8.873,11.24])
fx = vector([141.1,-35.51,167.1,18.3,173,243,1039,1783])
```

(leaving the rest untouched) and running the code results in nearly the same solution. It will differ slightly due to roundoff, but it is successful! Besides the simplicity of this method as compared to projection, being able to do the calculation using decimal approximations (floating point) is necessary for efficient computer implementation. Most programming languages do not provide exact computation, and those that do (like SageMath) are much slower. It is impractical to require exact arithmetic. Try it for yourself at SageMathCell 129.

3e: The coefficient matrix and constant vector for the linear regression problem are

$$M = \begin{bmatrix} x_1^2 & x_1t_1 & t_1^2 \\ x_2^2 & x_2t_1 & t_2^2 \\ \vdots & \vdots & \vdots \\ x_4^2 & x_4t_4 & t_4^2 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} -1.36 \\ 8.17 \\ \vdots \\ 50 \end{bmatrix}.$$

According to the code at SageMathCell 130 M is

```
[1681/1000000 12833/500000 97969/250000]
[23409/250000 47889/250000 97969/250000]
[ 11449/10000  33491/50000 97969/250000]
[    2304/625    3756/3125 97969/250000]
[1681/1000000  4141/100000  10201/10000]
[23409/250000  15453/50000  10201/10000]
[ 11449/10000  10807/10000  10201/10000]
[    2304/625     1212/625  10201/10000]
[1681/1000000  6273/100000  23409/10000]
[23409/250000  23409/50000  23409/10000]
```

```
[ 11449/10000  16371/10000  23409/10000]
[    2304/625     1836/625  23409/10000]
[1681/1000000  6683/100000  26569/10000]
[23409/250000  24939/50000  26569/10000]
[ 11449/10000  17441/10000  26569/10000]
[    2304/625     1956/625  26569/10000]
```

The code proceeds to calculate the normal equations $M^T M\hat{\mathbf{v}} = M^T\mathbf{b}$ and solve for $\hat{\mathbf{v}}$, the regression coefficients:

$$\begin{bmatrix} \beta_0 & \beta_1 & \beta_2 \end{bmatrix} = \begin{bmatrix} 4.88443881609 & 3.4469959515 & 8.28520093268 \end{bmatrix}^T$$

Hence the best fit model is

$$f(x) = 4.88443881609x^2 + 3.4469959515xt + 8.28520093268t^2.$$

Note: Unlike with the method of projection, it is not critical that the data are entered exactly nor that the computation is done using exact arithmetic. Replacing the first three lines of the code by

```
x = vector([.041,.306,1.07,1.92])
t = vector([.626,1.01,1.53,1.63])
ktx = vector([-1.36,8.17,6.59,24.1,8.47,8.88,16.9,38,18.9,
              19,29,46.1,21.4,28.9,36,50])
```

(leaving the rest untouched) and running the code results in nearly the same solution. It will differ slightly due to roundoff, but it is successful! Besides the simplicity of this method as compared to projection, being able to do the calculation using decimal approximations (floating point) is necessary for efficient computer implementation. Most programming languages do not provide exact computation, and those that do (like SageMath) are much slower. It is impractical to require exact arithmetic. Try it for yourself at SageMathCell 131.

4a: The coefficient matrix and constant vector for the linear regression problem are

$$M = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_8 & x_8^2 & x_8^3 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 141.1 \\ -35.51 \\ \vdots \\ 1783 \end{bmatrix}.$$

According to the code at SageMathCell 132 M is (with whitespace changes only)

```
[    1      389/1000       151321/1000000          58863869/1000000000]
[    1      851/1000       724201/1000000         616295051/1000000000]
[    1     2467/1000      6086089/1000000       15014381563/1000000000]
[    1     4113/1000     16916769/1000000       69578670897/1000000000]
[    1        181/40           32761/1600                5929741/64000]
[    1     6639/1000     44076321/1000000      292622695119/1000000000]
[    1     8873/1000     78730129/1000000      698572434617/1000000000]
[    1        281/25            78961/625               22188041/15625]
```

The code proceeds to orthogonalize the columns of M and then project $\mathbf{b}$ onto its column space. Using exact arithmetic as SageMath does, the projection is too long to print. It is approximately

$$\begin{bmatrix} 97.8 & 74.2 & 37.2 & 87.6 & 116 & 383 & 916 & 1816 \end{bmatrix}^T$$

This is the best approximation of $\mathbf{b}$ within the column space of M. Therefore, the regression coefficients (in the order $\beta_0, \beta_1, \beta_2, \beta_3$) are given by the solution of the system

$$M\hat{\mathbf{v}} = \begin{bmatrix} 97.8 & 74.2 & 37.2 & 87.6 & 116 & 383 & 916 & 1816 \end{bmatrix}^T$$

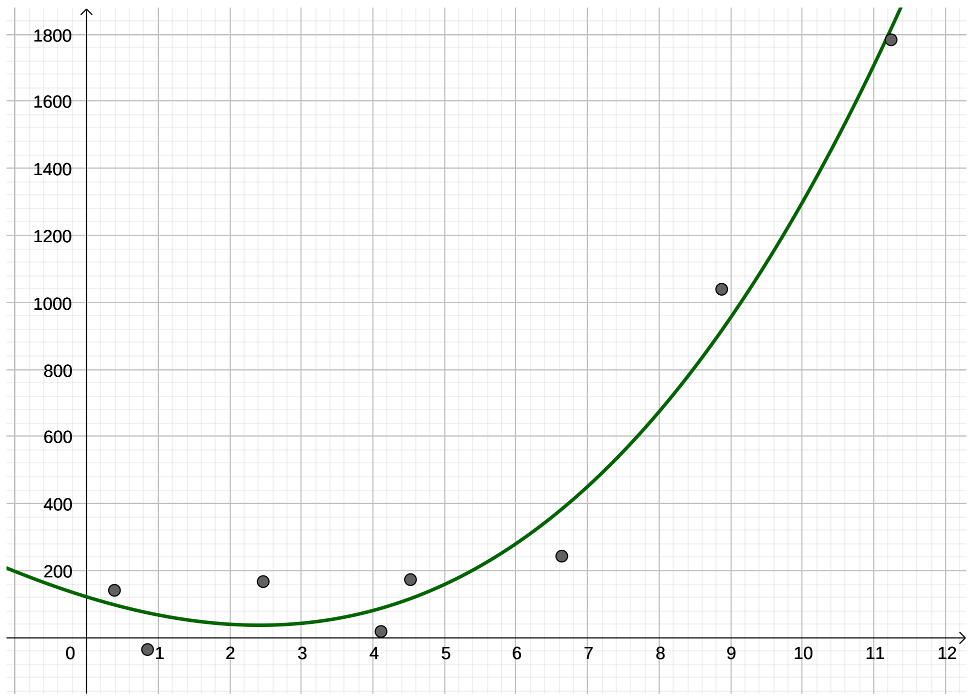
which according to the code (again approximately since the exact results are too long to print) is

$$\begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{bmatrix} = \begin{bmatrix} 121.697796 & -65.6594829 & 10.8873569 & 0.744289915 \end{bmatrix}^T$$

Hence the best fit model is

$$f(x) = 0.744289915x^3 + 10.8873569x^2 - 65.6594829x + 121.697796.$$

A plot of $f(x)$ superimposed on a scatterplot of the data is shown below, demonstrating geometrically the goodness of fit.



Note: It is critical that the data are entered exactly and that the computation is done using exact arithmetic. Replacing the first two lines of the code by

```
x = vector([.389,.851,2.467,4.113,4.525,6.639,8.873,11.24])
fx = vector([141.1,-35.51,167.1,18.3,173,243,1039,1783])
```

(leaving the rest untouched) and running the code results in an inconsistent system, due to roundoff error. The parameters are unsuccessfully calculated since the approximated projection of **b** is not in the column space of the approximated M. The rest of the calculation works perfectly well. Try it for yourself at SageMathCell 133.

4e: The coefficient matrix and constant vector for the linear regression problem are

$$M = \begin{bmatrix} x_1^2 & x_1t_1 & t_1^2 \\ x_2^2 & x_2t_1 & t_2^2 \\ \vdots & \vdots & \vdots \\ x_4^2 & x_4t_4 & t_4^2 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} -1.36 \\ 8.17 \\ \vdots \\ 50 \end{bmatrix}.$$

According to the code at SageMathCell 134 M is

```
[1681/1000000 12833/500000 97969/250000]
[23409/250000 47889/250000 97969/250000]
[ 11449/10000  33491/50000 97969/250000]
[    2304/625    3756/3125 97969/250000]
[1681/1000000  4141/100000  10201/10000]
[23409/250000  15453/50000  10201/10000]
[ 11449/10000  10807/10000  10201/10000]
[    2304/625     1212/625  10201/10000]
[1681/1000000  6273/100000  23409/10000]
[23409/250000  23409/50000  23409/10000]
[ 11449/10000  16371/10000  23409/10000]
[    2304/625     1836/625  23409/10000]
[1681/1000000  6683/100000  26569/10000]
[23409/250000  24939/50000  26569/10000]
[ 11449/10000  17441/10000  26569/10000]
[    2304/625     1956/625  26569/10000]
```

The code proceeds to orthogonalize the columns of M and then project $\mathbf{b}$ onto its column space. Using exact arithmetic as SageMath does, the projection is too long to print. It is approximately

$$\begin{bmatrix} 3.3 & 4.4 & 11 & 25 & 8.6 & 10 & 18 & 33 & 20 & 21 & 31 & 48 & 22 & 24 & 34 & 51 \end{bmatrix}^T$$

This is the best approximation of $\mathbf{b}$ within the column space of M. Therefore, the regression coefficients (in the order $\beta_0, \beta_1, \beta_2$) are given by the solution of the system

$$M\hat{\mathbf{v}} = \begin{bmatrix} 3.3 & 4.4 & 11 & 25 & 8.6 & 10 & 18 & 33 & 20 & 21 & 31 & 48 & 22 & 24 & 34 & 51 \end{bmatrix}^T$$

which according to the code (again approximately since the exact results are too long to print) is

$$\begin{bmatrix} \beta_0 & \beta_1 & \beta_2 \end{bmatrix} = \begin{bmatrix} 4.88443881609 & 3.4469959515 & 8.28520093268 \end{bmatrix}^T$$

Hence the best fit model is

$$f(x) = 4.88443881609x^2 + 3.4469959515xt + 8.28520093268t^2.$$

Note: It is critical that the data are entered exactly and that the computation is done using exact arithmetic. Replacing the first three lines of the code by

```
x = vector([.041,.306,1.07,1.92])
t = vector([.626,1.01,1.53,1.63])
ktx = vector([-1.36,8.17,6.59,24.1,8.47,8.88,16.9,38,18.9,
              19,29,46.1,21.4,28.9,36,50])
```

(leaving the rest untouched) and running the code results in an inconsistent system, due to roundoff error. The parameters are unsuccessfully calculated since the approximated projection of $\mathbf{b}$ is not in the column space of the approximated M. Try it for yourself at SageMathCell 135.

5a: The sum of the squared errors equals $||M\hat{\mathbf{v}} - \mathbf{b}||^2 = (M\hat{\mathbf{v}} - \mathbf{b}) \cdot (M\hat{\mathbf{v}} - \mathbf{b})$. Adding the lines

```
print(); print("Sum of squared errors:")
print((proj-fx)*(proj-fx))
```

to the code from the solution of question 3a (as seen at SageMathCell 136) gives a value of

$$||M\hat{\mathbf{v}} - \mathbf{b}||^2 \approx 74685.25$$

5e: The sum of the squared errors equals $||M\hat{\mathbf{v}} - \mathbf{b}||^2 = (M\hat{\mathbf{v}} - \mathbf{b}) \cdot (M\hat{\mathbf{v}} - \mathbf{b})$. Adding the lines

```
print(); print("Sum of squared errors:")
print((proj-ktx)*(proj-ktx))
```

to the code from the solution of question 3e (as seen at SageMathCell 137) gives a value of

$$\|M\hat{\mathbf{v}} - \mathbf{b}\|^2 \approx 125.160568$$

Section 7.2

1b: A transition matrix must be (i) square, (ii) have nonnegative entries, and (iii) have columns that sum to 1. The matrix

$$M = \begin{bmatrix} 0.09 & 0.686 & 0.168 \\ 0.908 & 0.036 & 0.807 \\ 0.002 & 0.278 & 0.485 \end{bmatrix}$$

has qualities (i) and (ii), but not (iii) since the third column does not sum to 1. Therefore, M is not a transition matrix.

Section 7.3

4b: In $C([0, L])$,

$$\left\langle \cos\left(m\frac{\pi}{L}t\right), \cos\left(m\frac{\pi}{L}t\right)\right\rangle = \int_0^L \cos^2\left(m\frac{\pi}{L}t\right)\,dt = \frac{1}{2}\int_0^L \left(1 + \cos\left(2m\frac{\pi}{L}t\right)\right)dt$$

$$= \frac{1}{2}\left[t + \frac{L}{2m\pi}\sin\left(2m\frac{\pi}{L}t\right)\right]_0^L dt = \frac{1}{2}\left[L + \frac{L}{2m\pi}\sin(2m\pi)\right] = \frac{L}{2}$$

Note that $\sin(2m\pi) = 0$ for any integer m.

5b: From question 4c, in $C([0, 1])$ $\langle \sin(m\pi t), \sin(m\pi t)\rangle = \frac{1}{2}$. Therefore

$$b_m = 2\langle f, \sin(m\pi t)\rangle = 2\int_0^1 t\sin(m\pi t)\,dt = \frac{-2}{m\pi}\left[t\cos(m\pi t) - \frac{1}{m\pi}\sin(m\pi t)\right]_0^1$$

$$\frac{-2}{m\pi}\cos(m\pi) = \frac{-2}{m\pi}(-1)^m = \frac{2}{m\pi}(-1)^{m+1}$$

and the Fourier sine series is

$$\frac{2}{\pi}\sin(\pi t) - \frac{1}{\pi}\sin(2\pi t) + \frac{2}{3\pi}\sin(3\pi t) - \cdots$$

Notes:

1. $\sin(m\pi) = 0$ for any integer m
2. $\int t\sin(m\pi t)\,dt$ can be calculated using integration by parts:

$$\int t\sin(m\pi t)\,dt = -\frac{t}{m\pi}\cos(m\pi t) + \frac{1}{m\pi}\int \cos(m\pi t)\,dt$$

$$= -\frac{t}{m\pi}\cos(m\pi t) + \frac{1}{(m\pi)^2}\sin(m\pi t) = \frac{-1}{m\pi}\left[t\cos(m\pi t) - \frac{1}{m\pi}\sin(m\pi t)\right]$$

6b: From question 4c, in $C([0, 1])$ $\langle \cos(m\pi t), \cos(m\pi t)\rangle = \frac{1}{2}$ and from question 4a, $\langle 1, 1\rangle = 1$. Therefore,

$$a_0 = \langle f, 1\rangle = \int_0^1 t\,dt = \left[\frac{1}{2}t^2\right]_0^1 = \frac{1}{2}$$

and for $m = 1, 2, \ldots$

$$a_m = 2\langle f, \cos(m\pi t)\rangle = 2\int_0^1 t\cos(m\pi t)\ dt = \frac{2}{m\pi}\left[t\sin(m\pi t) + \frac{1}{m\pi}\cos(m\pi t)\right]_0^1$$

$$\frac{2}{(m\pi)^2}\left[\cos(m\pi) - 1\right] = \frac{2}{(m\pi)^2}\left[(-1)^m - 1\right] = \frac{-4}{(m\pi)^2} \text{ for } m = 1, 3, 5, \ldots$$

and the Fourier cosine series is

$$\frac{1}{2} - \frac{4}{\pi^2}\cos(\pi t) - \frac{4}{9\pi^2}\cos(3\pi t) - \frac{4}{25\pi^2}\cos(5\pi t) - \cdots$$

Notes:

1. $\sin(m\pi) = 0$ for any integer m
2. $\int t\cos(m\pi t)\ dt$ can be calculated using integration by parts:

$$\int t\cos(m\pi t)\ dt = \frac{t}{m\pi}\sin(m\pi t) - \frac{1}{m\pi}\int \sin(m\pi t)\, dt$$

$$= \frac{t}{m\pi}\sin(m\pi t) + \frac{1}{(m\pi)^2}\cos(m\pi t) = \frac{1}{m\pi}\left[t\sin(m\pi t) + \frac{1}{m\pi}\cos(m\pi t)\right]$$

7a: From question 4, in $C([0,1])$, $\langle\cos(m\pi t), \cos(m\pi t)\rangle = \langle\sin(m\pi t), \sin(m\pi t)\rangle = \frac{1}{2}$ and $\langle 1, 1\rangle = 1$. Combined with the fact that the functions 1, $\cos^2(m\pi t)$, and $\sin^2(m\pi t)$ are even, in $C([-1,1])$ $\langle\cos(m\pi t), \cos(m\pi t)\rangle = \langle\sin(m\pi t), \sin(m\pi t)\rangle = 1$ and $\langle 1, 1\rangle = 2$. Therefore,

$$a_0 = \frac{1}{2}\langle f, 1\rangle = \frac{1}{2}\int_{-1}^1 1\, dt = \frac{1}{2}\left[t\right]_{-1}^1 = 1$$

and for $m = 1, 2, \ldots$

$$a_m = \langle f, \cos(m\pi t)\rangle = \int_{-1}^1 \cos(m\pi t)\ dt = \frac{2}{m\pi}\left[\sin(m\pi t)\right]_{-1}^1 = 0$$

$$b_m = \langle f, \sin(m\pi t)\rangle = \int_{-1}^1 \sin(m\pi t)\ dt = 0 \text{ (because } \sin(m\pi t) \text{ is odd)}$$

and the Fourier cosine series is

$$1$$

Section 7.4

1a: Though SageMath is not strictly required, it makes repetitive calculations like this more manageable.

$$\mathbf{x}_1 = \mathbf{f}(\mathbf{x}_0) = \frac{1}{42}\begin{bmatrix} 16 & 26 \\ 65 & -23 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1/7 \\ 3/14 \end{bmatrix} = \begin{bmatrix} 1/7 \\ 3/14 \end{bmatrix} \approx \begin{bmatrix} 0.14285 \\ 0.21428 \end{bmatrix},$$

$$\mathbf{x}_2 = \mathbf{f}(\mathbf{x}_1) = \frac{1}{42}\begin{bmatrix} 16 & 26 \\ 65 & -23 \end{bmatrix}\begin{bmatrix} 1/7 \\ 3/14 \end{bmatrix} + \begin{bmatrix} 1/7 \\ 3/14 \end{bmatrix} = \begin{bmatrix} 55/294 \\ 61/588 \end{bmatrix} \approx \begin{bmatrix} 0.18707 \\ 0.10374 \end{bmatrix},$$

$$\mathbf{x}_3 = \mathbf{f}(\mathbf{x}_2) = \frac{1}{42}\begin{bmatrix} 16 & 26 \\ 65 & -23 \end{bmatrix}\begin{bmatrix} 55/294 \\ 61/588 \end{bmatrix} + \begin{bmatrix} 1/7 \\ 3/14 \end{bmatrix} = \begin{bmatrix} 239/1764 \\ 821/3528 \end{bmatrix} \approx \begin{bmatrix} 0.13548 \\ 0.23270 \end{bmatrix},$$

$$\mathbf{x}_4 = \mathbf{f}(\mathbf{x}_3) = \frac{1}{42}\begin{bmatrix} 16 & 26 \\ 65 & -23 \end{bmatrix}\begin{bmatrix} 239/1764 \\ 821/3528 \end{bmatrix} + \begin{bmatrix} 1/7 \\ 3/14 \end{bmatrix} = \begin{bmatrix} 2071/10584 \\ 1741/21168 \end{bmatrix} \approx \begin{bmatrix} 0.19567 \\ 0.08224 \end{bmatrix}$$

2a: A fixed point of a function $f(x)$ is a solution of the equation $f(x) = x$ (where the input and output of the function are equal). Iteration beginning at the fixed point stays at the fixed point. Solving for the fixed points:

$$\frac{1}{42}\begin{bmatrix} 16 & 26 \\ 65 & -23 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1/7 \\ 3/14 \end{bmatrix} = \mathbf{x}$$

$$\frac{1}{42}\begin{bmatrix} 16 & 26 \\ 65 & -23 \end{bmatrix}\mathbf{x} - \mathbf{x} = -\begin{bmatrix} 1/7 \\ 3/14 \end{bmatrix}$$

$$\left(\frac{1}{42}\begin{bmatrix} 16 & 26 \\ 65 & -23 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)\mathbf{x} = -\begin{bmatrix} 1/7 \\ 3/14 \end{bmatrix}$$

$$\frac{1}{42}\begin{bmatrix} -26 & 26 \\ 65 & -65 \end{bmatrix}\mathbf{x} = -\begin{bmatrix} 1/7 \\ 3/14 \end{bmatrix}$$

so fixed points are solutions of

$$\begin{bmatrix} -26 & 26 \\ 65 & -65 \end{bmatrix}\mathbf{x} = \begin{bmatrix} -6 \\ -9 \end{bmatrix},$$

an inconsistent system. **f** has no fixed points.

4a: The dynamical system has no fixed point, so it cannot have an attractor (an attractor is a fixed point).

5a: The solution of question 1a indicates the orbit is not constant. The answer to question 3a gives eigenvalues of $-\frac{7}{6}$ and 1 so the spectral radius of $\frac{1}{42}\begin{bmatrix} 16 & 26 \\ 65 & -23 \end{bmatrix}$ is greater than one $\left(\frac{7}{6}\text{ in this case}\right)$. With a spectral radius greater than one and an orbit that is not fixed, the dyamical system will tend toward infinity $\left(\text{at approximately the rate } \left(\frac{7}{6}\right)^k\right)$.

5j: With a spectral radius less than one $\left(\frac{1}{2}\text{ in this case}\right)$, the fixed point will be an attractor, and the dynamical system will tend toward the fixed point.

Section 7.5

1b: Answers may vary. The figure is scaled by $\frac{1}{2}$ in both the horizontal and vertical directions, either reflected about the y-axis or rotated 90°, and then translated into place. To be more specific, these general observations lead to two possibilities (and there are others):

1. scale by $\frac{1}{2}$, reflect about the y-axis, translate $\begin{bmatrix} 4 \\ 0 \end{bmatrix}$. As a formula,

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

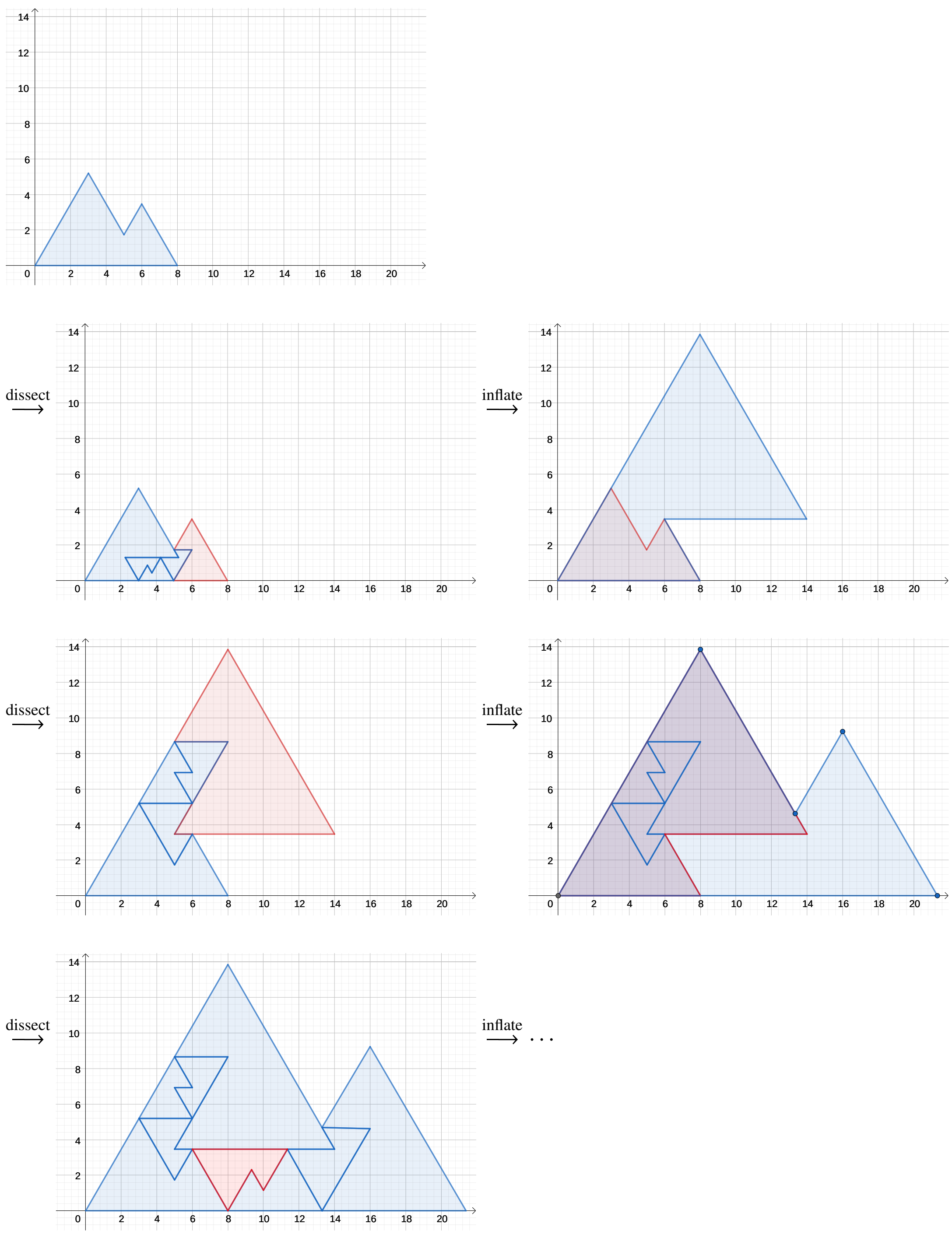
$$= \begin{bmatrix} -1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

2. scale by $\frac{1}{2}$, rotate 90° (counterclockwise) about the origin, translate $\begin{bmatrix} 4 \\ 0 \end{bmatrix}$. As a formula,

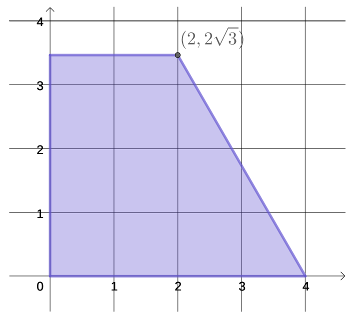
$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -1/2 \\ 1/2 & 0 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

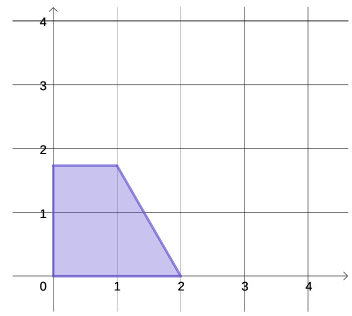
3: One of the easiest ways to show that a shape tessellates the plane is to show that it tiles a parallelogram or a triangle. That tiling can then be repeated to tessellate the plane. In this case, however, that is not an option. A dissect and inflate, dissect and inflate, dissect and inflate,... procedure can be used instead. The inflation is performed so that one part of the dissection becomes the whole (and the other parts lie outside the whole). Which part becomes the whole rotates among the parts. In pictures:

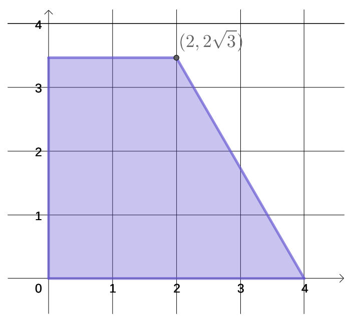


9d: The IFS contains an affine transformation for each part of the dissection. Each transformation maps the whole to one of the parts.

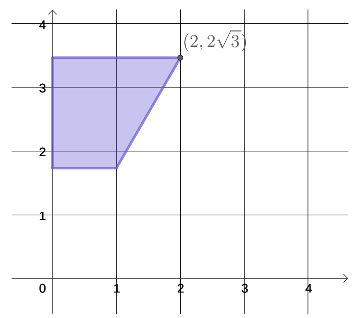


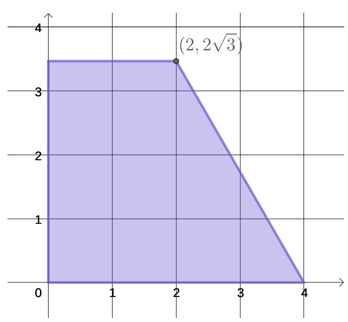
scale $\frac{1}{2}$
$\longrightarrow$



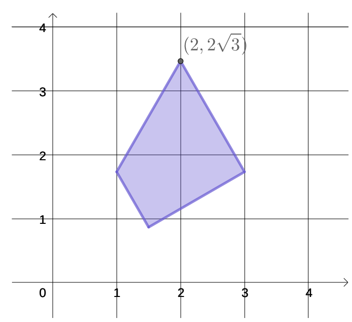


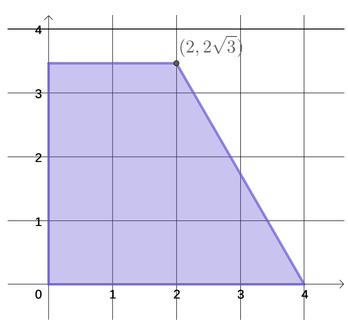
scale $\frac{1}{2}$
$\longrightarrow$
reflect about x-axis
translate $(0, 2\sqrt{3})$



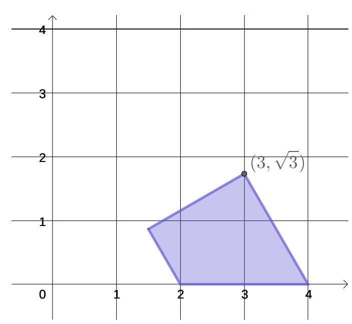


scale $\frac{1}{2}$
$\longrightarrow$
rotate $120°$
translate $(3, \sqrt{3})$





scale $\frac{1}{2}$
$\longrightarrow$
reflect about x-axis
rotate $-60°$
translate $(3, \sqrt{3})$



Formally, the IFS is

$$\left\{T_1(\mathbf{x}) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{x}, \quad T_2(\mathbf{x}) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 2\sqrt{3} \end{bmatrix},\right.$$

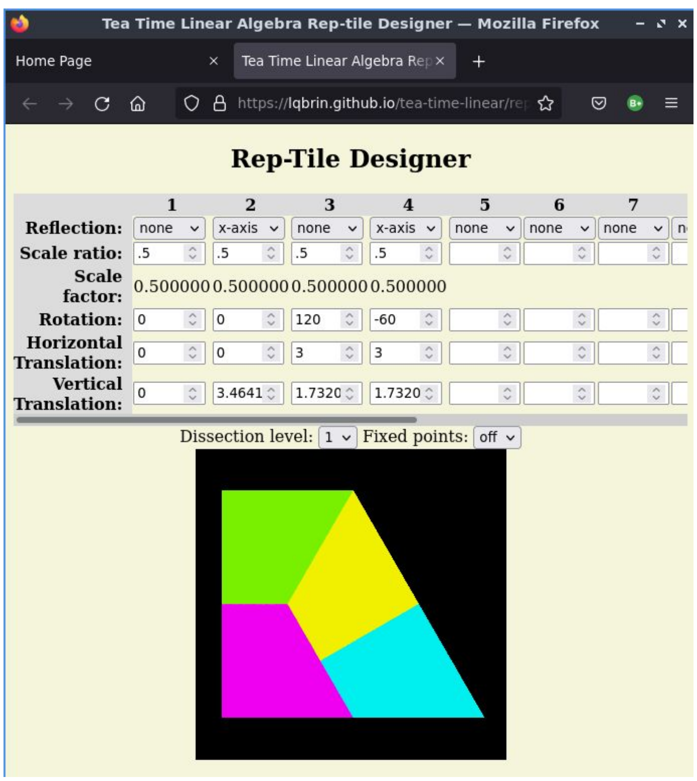
$$\left. T_3(\mathbf{x}) = \begin{bmatrix} -\frac{1}{4} & -\frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & -\frac{1}{4} \end{bmatrix}\mathbf{x} + \begin{bmatrix} 3 \\ \sqrt{3} \end{bmatrix}, \quad T_4(\mathbf{x}) = \begin{bmatrix} \frac{1}{4} & \frac{\sqrt{3}}{4} \\ -\frac{\sqrt{3}}{4} & \frac{1}{4} \end{bmatrix}\mathbf{x} + \begin{bmatrix} 3 \\ \sqrt{3} \end{bmatrix}\right\}$$

Note: There are different possible literal descriptions of the transformations, but the IFS itself is unique.

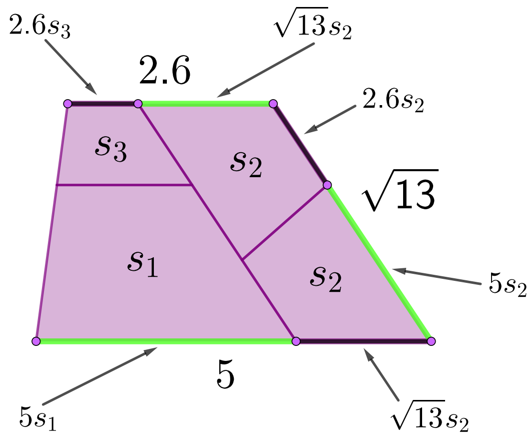
10: Taking the literal descriptions from the solution of exercise 9d and translating them to the required sequence:

scale $\frac{1}{2}$ $\longrightarrow$	none, $\frac{1}{2}$, 0, 0, 0
scale $\frac{1}{2}$ $\longrightarrow$ reflect about x-axis translate $(0, 2\sqrt{3})$	x-axis, $\frac{1}{2}$, 0, 0, $2\sqrt{3} \approx 3.46410$
scale $\frac{1}{2}$ $\longrightarrow$ rotate 120° translate $(3, \sqrt{3})$	none, $\frac{1}{2}$, 120, 3, $\sqrt{3} \approx 1.73205$
scale $\frac{1}{2}$ $\longrightarrow$ reflect about x-axis rotate $-60°$ translate $(3, \sqrt{3})$	x-axis, $\frac{1}{2}$, -60, 3, $\sqrt{3} \approx 1.73205$

A screenshot of this information in the rep-tile designer shows the correct rep-tile, verifying that the transformations are correct.



11: Each part is labeled with its scale factor in the following diagram. Each of the labeled sides generates one equation involving the scale factors by the fact that the lengths of the two parts of each side must sum to the length of the whole side.



$$\begin{aligned} 5s_1 + \sqrt{13}s_2 &= 5 \\ 2.6s_2 + 5s_2 &= \sqrt{13} \\ 2.6s_3 + \sqrt{13}s_2 &= 2.6 \end{aligned}$$

In matrix form, the system of equations is

$$\begin{bmatrix} 5 & \sqrt{13} & 5 \\ 0 & \frac{38}{5} & \sqrt{13} \\ 0 & \sqrt{13} & \frac{13}{5} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} 5 \\ \sqrt{13} \\ \frac{13}{5} \end{bmatrix}$$

and can be solved using the code at SageMathCell 138. The solution is $s_1 = \frac{25}{38}$, $s_2 = \frac{5\sqrt{13}}{38}$, $s_3 = \frac{13}{38}$. Note: the code also verifies that the sum of the squares of the scale factors $\left(s_1^2 + 2s_2^2 + s_3^2\right)$ is 1, evidence that the solution is correct.

Answers to Selected Exercises

Section 1.1

5b: 1×6

5f: 2×2

6d: $\begin{bmatrix} -1 \\ 1 \\ 0 \\ -8 \\ -5 \\ -3 \end{bmatrix}$

6h: $\begin{bmatrix} -7 & 2 & 1 & 8 \\ -8 & -11 & 10 & -6 \\ 9 & -1 & 3 & -6 \\ 6 & -9 & 3 & 4 \\ 2 & -4 & 10 & -7 \end{bmatrix}$

9: (a) `C = matrix(3,2,[-4,-9,13,-11,7,5,-14,-2,-12,12,8,11])` (b) `print(C)` (c) `entry = C[3,2]` (d) `print(entry)`

Section 1.2

1c: $\begin{bmatrix} -2 & -12 & 14 & -3 \\ -8 & -5 & 3 & -5 \end{bmatrix}$

4: Yes. Explain.

Section 1.3

1f: -23

1g: 15.26

2g: $\begin{bmatrix} 11 & 6 \\ 11 & 6 \end{bmatrix}$

2k: $\begin{bmatrix} 18 & 9 & 3 \\ 6 & 11 & -23 \end{bmatrix}$

2m: undefined

3c: 27

3g: 77

3j: -43

4c: 27

4g: 77

4j: -43

5: 2×5

8: (a) true (b) true

Section 1.4

1c: $\sqrt{109}$

1h: $\sqrt{30.42} \approx 5.51543289325507$

1j: $3\sqrt{15}$

2c: $2\sqrt{34}$

2h: $\sqrt{6.81} \approx 2.60959767013998$

2j: $\sqrt{91}$

3c: yes

3h: no

3j: yes

4f: $k = -14,\ 6$

Section 1.5

1d: $C_{2,2} = \det(8)$

1f: $C_{2,1} = -\det\begin{pmatrix} 0 & 1 \\ 11 & -7 \end{pmatrix}$

2d: 44

2h: 97

2l: 273

4: (a) true (b) false (c) false (d) true (e) false

10: It is not possible to write any row as a linear combination of the other two in any of the remaining seven blocks.

Section 1.6

1d: $\begin{bmatrix} \frac{1}{4\pi} \end{bmatrix}$

1g: $\begin{bmatrix} 1 & -\sqrt{2} \\ -\frac{2\sqrt{2}}{3} & \frac{5}{3} \end{bmatrix}$

1n: $\begin{bmatrix} 1 & 2 & 0 \\ -2 & 6 & -3 \\ 1 & -1 & 1 \end{bmatrix}$

1p: Not possible.

1q: $\begin{bmatrix} 13 & -8 & 5 & 1 \\ 5 & 0 & 4 & 2 \\ 6 & -3 & 3 & 1 \\ 3 & -1 & 2 & 1 \end{bmatrix}$

Section 1.7

1b: -3

1f: 8

2c: $\lambda^2 + 21\lambda$

2e: $\lambda^3 - 2\lambda^2 - \lambda + 2$

3b: $-2, 3$

3f: $-1, 0$

3j: $-2 \pm i\sqrt{2}$

3k: $-1, 0, 1$

4b: any vector of the form $r\begin{bmatrix} 5 \\ 1 \end{bmatrix}$.

4e: any vector of the form $r\begin{bmatrix} 4 \\ -3 - i\sqrt{3} \end{bmatrix}$.

4h: any vector of the form $r\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$.

6: (a) false (b) false (c) true (d) false (e) false (f) true (g) false

9: The eigenvector is $\begin{bmatrix} -1 \\ 2 \\ -6 \\ -2 \end{bmatrix}$ and the associated eigenvalue is 9.

Section 2.1

1b: $\begin{bmatrix} 3 & 2 & -8 & 9 \\ 0 & -3 & 2 & 10 \\ -7 & 1 & 0 & -11 \end{bmatrix}$

1d: $\begin{bmatrix} 3 & 2 & -3 & -7 \\ -5 & 1 & -2 & -8 \\ 1 & 1 & 1 & 11 \end{bmatrix}$

2b: $\begin{array}{rcrcrcr} -14v_1 & - & 15v_2 & + & 8v_3 & = & -8 \\ -13v_1 & + & 2v_2 & - & v_3 & = & 13 \\ 15v_1 & - & 9v_2 & - & 6v_3 & = & 12 \end{array}$

2d:

$$10v_1 - 9v_2 - v_3 + 3v_4 + 15v_5 = 6$$
$$-11v_1 + 12v_2 + 13v_3 + 5v_4 - 4v_5 = -2$$

5: One solution is $v_1 = -5$, $v_2 = 4$, $v_3 = 0$. It has infinitely many more because both v_1 and v_2 can be written in terms of v_3, and v_3 can take any value:

$$v_1 = -5 - 2v_3 \quad \text{and} \quad v_2 = 4 - 3v_3$$

The example solution comes from taking $v_3 = 0$.

6b: Swap rows 2 and 3

6d: Row 2 replaced by row 2 minus 9 row 3

8b: $\begin{bmatrix} 3 & -1 & 1 & -7 \\ 5 & -10 & 30 & -20 \end{bmatrix}$

8d: $\begin{bmatrix} -9 & 9 & -4 \\ 5 & 6 & -2 \\ -9 & -9 & -6 \end{bmatrix}$

8f: $\begin{bmatrix} -5 & 1 & 9 & -6 \\ 22 & -10 & -44 & 29 \\ -8 & 3 & 2 & 2 \end{bmatrix}$

9d: Swap rows 1 and 3

9e: Scale row 3 by 3

10b: Replace row 1 by row 1 minus 2 row 2:

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

11b: Replace row 2 by row 2 minus 2 row 1:

$$\begin{bmatrix} 1 & 0 & -3 \\ 0 & 3 & 4 \end{bmatrix}$$

14: Each coefficient of the linear combination that gives the second row of AB is the same multiple of the corresponding coefficient of the linear combination that gives the first row of AB, so the second row must be that multiple of the first row:

$$\begin{aligned} (AB)_{1,:} &= A_{1,1}B_{1,:} + A_{1,2}B_{2,:} + \cdots + A_{1,n}B_{n,:} \\ (AB)_{2,:} &= kA_{1,1}B_{1,:} + kA_{1,2}B_{2,:} + \cdots + kA_{1,n}B_{n,:} \\ &= k\left(A_{1,1}B_{1,:} + A_{1,2}B_{2,:} + \cdots + A_{1,n}B_{n,:}\right) \\ &= k(AB)_{1,:} \end{aligned}$$

Section 2.2

3g: Answers will vary. Any answer where $x_3 = \frac{4}{3}x_4$ and $x_1 = -\frac{3}{5}x_2$ (but not all zero) is correct. For example, $x_1 = -3$, $x_2 = 5$, $x_3 = 4$, $x_4 = 3$ is one solution.

4d: $x = -6$, $y = -\frac{21}{5}$

4f: $x = \frac{22}{5}$, $y = -24$

4j: $x_1 = -\frac{4}{3}$, $x_2 = -\frac{1}{3}$, $x_3 = -\frac{8}{7}$

4l: $x_1 = -92$, $x_2 = -42$, $x_3 = 32$, $x_4 = 77$

Section 2.3

1b: (i) yes (ii) consistent (iii) 1

1c: (i) no

1d: (i) yes (ii) inconsistent (iii) 0

1h: (i) yes (ii) consistent (iii) infinitely many

2d: (i) not in reduced row echelon form

2f: (i) in reduced row echelon form

(ii) $\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 0 \\ 2 \end{bmatrix} + r \begin{bmatrix} 9 \\ -1 \\ 1 \\ 0 \end{bmatrix}$

3c: (i) yes (ii) $\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = r \begin{bmatrix} 1 \\ -\frac{1}{3} \\ 1 \end{bmatrix}$

3d: (i) no

4c: $\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

4f: $\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = r \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} 6 \\ 0 \\ 1 \end{bmatrix}$

5b: $r \begin{bmatrix} 1 \\ -\frac{1}{2} \end{bmatrix}, s \begin{bmatrix} 1 \\ -\frac{2}{3} \end{bmatrix}$

10: (a) yes, yes (b) infinitely many, infinitely many (c) 1, 1 (d) no (e) yes

12: The coefficient matrix will have more columns than rows, and therefore the system will have at least one free variable.

13: Yes. For example

$$\begin{array}{rcrcr} 3x & + & 2y & = & 7 \\ x & - & 3y & = & -5 \\ 5x & - & 4y & = & -3 \end{array}$$

has solution $x = 1$, $y = 2$.

Section 3.1

1e: Answers will vary—examples will be different as they are creations of individuals, and explanations will be different as they are requested as informal explanations, likely based on intuition rather than definition. Let $r = 3$, $s = 2$ and $A = \begin{bmatrix} -1 & 2 \\ 4 & -5 \end{bmatrix}$. Then $(rs)A = (2 \cdot 3)A = 6A = \begin{bmatrix} 6 \cdot -1 & 6 \cdot 2 \\ 6 \cdot 4 & 6 \cdot -5 \end{bmatrix} = \begin{bmatrix} -6 & 12 \\ 24 & -30 \end{bmatrix}$ and $r(sA) = 3(2A) = 3\left(\begin{bmatrix} 2 \cdot -1 & 2 \cdot 2 \\ 2 \cdot 4 & 2 \cdot -5 \end{bmatrix}\right) = 3\begin{bmatrix} -2 & 4 \\ 8 & -10 \end{bmatrix} = \begin{bmatrix} 3 \cdot -2 & 3 \cdot 4 \\ 3 \cdot 8 & 3 \cdot -10 \end{bmatrix} = \begin{bmatrix} -6 & 12 \\ 24 & -30 \end{bmatrix}$. The rule is true essentially because the associative rule for multiplication of real numbers applies to each entry.

1i: Answers will vary—examples will be different as they are creations of individuals, and explanations will be different as they are requested as informal explanations, likely based on intuition rather than definition. Let $A = \begin{bmatrix} 3 & -4 & -9 \\ 7 & -6 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 9 & 1 \\ 4 & 7 & -2 \end{bmatrix}$. Then $(A+B)^T = \left(\begin{bmatrix} 3 & -4 & -9 \\ 7 & -6 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 9 & 1 \\ 4 & 7 & -2 \end{bmatrix}\right)^T =$

$\begin{bmatrix} 4 & 5 & -8 \\ 11 & 1 & 0 \end{bmatrix}^T = \begin{bmatrix} 4 & 11 \\ 5 & 1 \\ -8 & 0 \end{bmatrix}$ while $A^T + B^T = \begin{bmatrix} 3 & -4 & -9 \\ 7 & -6 & 2 \end{bmatrix}^T + \begin{bmatrix} 1 & 9 & 1 \\ 4 & 7 & -2 \end{bmatrix}^T = \begin{bmatrix} 3 & 7 \\ -4 & -6 \\ -9 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 4 \\ 9 & 7 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 4 & 11 \\ 5 & 1 \\ -8 & 0 \end{bmatrix}$. The rule is true because corresponding entries of A and B are also corresponding entries of A^T and B^T. Thus the two numbers being added to obtain each entry of the sum is the same on either side of the equation.

5: (a) $A^{-1} = \begin{bmatrix} -5 & 7 \\ 3 & -4 \end{bmatrix}$ (b) Answers will vary. $(A^T)^{-1} = \begin{bmatrix} -5 & 3 \\ 7 & -4 \end{bmatrix}$ because the inverse of the transpose is the transpose of the inverse (theorem 4 claim 8)

16b: Answers will vary. One demonstration relies on the fact that for any matrix M, $M - M = 0$ (justified in an answer on page 77): $A + (-A) = A + (-1 \cdot A) = A - A = 0$. Another demonstration relies on distributivity (theorem 3 claim 2): $A + (-A) = 1 \cdot A + (-1 \cdot A) = (1 + (-1))A = 0A = 0$.

17: Answers will vary. $(B^{-1}A^{-1})(AB) = ((B^{-1}A^{-1})A)B = (B^{-1}(A^{-1}A))B = (B^{-1}I)B = B^{-1}B = I$ using the associative property multiple times, the definition of inverse multiple times, and the definition of the multiplicative identity once. The second half of the justification, that $(AB)(B^{-1}A^{-1}) = I$, can be made by a similar string of equalities. You are encouraged to try it.

Section 3.2

1b: $\begin{bmatrix} 16 & -13 \\ -2 & -3 \end{bmatrix}$

2c: $\begin{bmatrix} 5 & -14 \\ 62 & -136 \end{bmatrix}$

3b: $Z = Y^{-1}X^{-1}B$ assuming X and Y are invertible

3e: $C = \frac{1}{2}D^T(A^{-1})^T - (B^{-1})^T$ assuming A and B are invertible

4c:
$$\begin{aligned} 10r + 14s &= -8 \\ -6r - 3s &= 9 \end{aligned}$$

5d: $M = \begin{bmatrix} 14 & -17 \\ 2 & 0 \\ 9 & -2 \end{bmatrix}$; $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} -11 \\ -4 \\ -8 \end{bmatrix}$

6d: $\begin{bmatrix} 11 & 12 & -9 \end{bmatrix}$

7d: The product has no third row

8d: $\begin{bmatrix} -14 \\ 12 \end{bmatrix}$

9d: $\begin{bmatrix} 16 \\ -9 \end{bmatrix}$

Section 3.3

8: linearly independent

12: (a) must have 3 pivots: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ (b) must have 2,1, or 0 pivots: $\begin{bmatrix} 1 & 0 & \star \\ 0 & 1 & \star \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & \star & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

15b: $x \neq -\frac{8}{7}$

15d: $x \neq \frac{26}{3}$

16b: $x = \frac{10}{7}$

16d: $x = -8$

22b: (i) $7R_{:,1} + 9R_{:,2} = 0$ (ii) $7\begin{bmatrix} -9 \\ 0 \\ 0 \\ 0 \end{bmatrix} + 9\begin{bmatrix} 7 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ (iii) $7\begin{bmatrix} 180 \\ -108 \\ -72 \\ -189 \end{bmatrix} + 9\begin{bmatrix} -140 \\ 84 \\ 56 \\ 147 \end{bmatrix} = \begin{bmatrix} 1260 \\ -756 \\ -504 \\ -1323 \end{bmatrix} + \begin{bmatrix} -1260 \\ 756 \\ 504 \\ 1323 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

23b: (i) columns 1 and 3 are linearly independent; and columns 2 and 4 are linearly independent. (ii) columns 1 and 3 are not multiples of one another and therefore are linearly independent; and columns 2 and 4 are not multiples of one another and therefore are linearly independent. (iii) columns 1 and 3 are not multiples of one another and therefore are linearly independent; and columns 2 and 4 are not multiples of one another and therefore are linearly independent.

Section 3.4

1d: (i) no (ii) yes (iii) one (iv) one

2c: (i) 9 (ii) yes (iii) no (iv) infinitely many (v) infinitely many

3c: (ii) no (iii) no (iv) infinitely many (v) infinitely many

7: 5

10: (a) must have three pivots:

$$\begin{bmatrix} 1 & 0 & 0 & \star \\ 0 & 1 & 0 & \star \\ 0 & 0 & 1 & \star \end{bmatrix}, \begin{bmatrix} 1 & 0 & \star & 0 \\ 0 & 1 & \star & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$
$$\begin{bmatrix} 1 & \star & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(b) must have 2,1, or 0 pivots:

$$\begin{bmatrix} 1 & 0 & \star & \star \\ 0 & 1 & \star & \star \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & \star & 0 & \star \\ 0 & 0 & 1 & \star \\ 0 & 0 & 0 & 0 \end{bmatrix},$$
$$\begin{bmatrix} 1 & \star & \star & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 & \star \\ 0 & 0 & 1 & \star \\ 0 & 0 & 0 & 0 \end{bmatrix},$$
$$\begin{bmatrix} 0 & 1 & \star & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$
$$\begin{bmatrix} 1 & \star & \star & \star \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & \star & \star \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$
$$\begin{bmatrix} 0 & 0 & 1 & \star \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$
$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

12b: 82

14b: $x \neq -\frac{8}{7}$; same as section 3.3 exercise 15b

14d: $x \neq \frac{26}{3}$; same as section 3.3 exercise 15d

15b: $x = \frac{10}{7}$; same as section 3.3 exercise 16b

15d: $x = -8$; same as section 3.3 exercise 16d

Section 3.5

5: (a) -38 (b) 19 (c) 38 (d) -190

8: 2

11: $\frac{1}{3} \cdot 7^n$

14: $-\frac{1}{2}$

Section 3.6

11: Theorem 7 part (ix) if false for G so part (v) is also false, from which we conclude $G\mathbf{v} = \mathbf{b}$ has infinitely many solutions (in addition to the trivial solution).

13: Theorem 7 part (xiii) is true for M so part (i) is as well, from which we conclude that the columns of M are linearly independent. Since the rows of M^T are the same as the columns of M, the rows of M^T are linearly independent.

15: Theorem 7 part (x) is false for B so (a) its columns are linearly dependent [by theorem 7 part (i)] (b) $B\mathbf{v} = \mathbf{0}$ has infinitely many solutions [by theorem 7 part (v)] (c) there is no matrix A such that $AB = I$ [by theorem 7 part (xi)]

17: (a) false (b) false (c) false (d) false (e) true (f) false (g) true

Section 3.7

8: (a) No. A need not be square (it may be 2×3 for example) and therefore not invertible. (b) Yes. $\det(AA^T) = (\det A)(\det A^T) = (\det A)^2 = \det\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} = 5$ so $\det A = \sqrt{5} \neq 0$.

10: No. By theorem 7 $M - cI$ must be invertible and therefore $\det(M - cI) \neq 0$.

11c: $A - \lambda I = \begin{bmatrix} 3 + i\sqrt{3} & 4 \\ -3 & -3 + i\sqrt{3} \end{bmatrix}$ reduces to $\begin{bmatrix} 3 + i\sqrt{3} & 4 \\ 0 & 0 \end{bmatrix}$ so an eigenvector $\mathbf{v}$ must satisfy $v_2 = -\frac{3+i\sqrt{3}}{4} v_1$. In parametric vector form,

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ -\frac{3+i\sqrt{3}}{4} v_1 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ -\frac{3+i\sqrt{3}}{4} \end{bmatrix}.$$

Using generic free variables (and multiplying by -4):

$$\mathbf{v} = r \begin{bmatrix} -4 \\ 3 + i\sqrt{3} \end{bmatrix}$$

11e: $A - \lambda I = \begin{bmatrix} -12 & 9 & 18 \\ 12 & -9 & -18 \\ 12 & -9 & -18 \end{bmatrix}$ reduces to $\begin{bmatrix} -4 & 3 & 6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ so an eigenvector $\mathbf{v}$ must satisfy $v_1 = \frac{3}{4}v_2 + \frac{3}{2}v_3$. In parametric vector form,

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} \frac{3}{4}v_2 + \frac{3}{2}v_3 \\ v_2 \\ v_3 \end{bmatrix} = v_2 \begin{bmatrix} \frac{3}{4} \\ 1 \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} \frac{3}{2} \\ 0 \\ 1 \end{bmatrix}.$$

Using generic free variables (and scaling to eliminate fractions):

$$\mathbf{v} = r \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} + s \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}$$

11g: $A - \lambda I = \begin{bmatrix} -45 & -51 & -24 & -60 \\ 15 & 17 & 18 & 0 \\ 15 & 17 & 8 & 20 \\ -30 & -34 & -16 & -40 \end{bmatrix}$ reduces to $\begin{bmatrix} 15 & 17 & 0 & 36 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ so an eigenvector $\mathbf{v}$ must satisfy $v_1 = -\frac{17}{15}v_2 - \frac{36}{15}v_4$, $v_3 = 2v_4$. In parametric vector form,

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} -\frac{17}{15}v_2 - \frac{36}{15}v_4 \\ v_2 \\ 2v_4 \\ v_4 \end{bmatrix} = v_2 \begin{bmatrix} -\frac{17}{15} \\ 1 \\ 0 \\ 0 \end{bmatrix} + v_4 \begin{bmatrix} -\frac{36}{15} \\ 0 \\ 2 \\ 1 \end{bmatrix}.$$

Using generic free variables (and scaling to eliminate fractions):

$$\mathbf{v} = r \begin{bmatrix} -17 \\ 15 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -36 \\ 0 \\ 30 \\ 15 \end{bmatrix}$$

Section 4.1

4d: The line passing through $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$

4i: The xy-plane

5a: The set of all sequences whose terms are all equal except possibly the third, which is twice the others

5d: All constant functions (those whose graphs are horizontal lines)

12c: No

12f: Yes

14a: Yes

14f: Yes

15c: Yes

21: Yes

16c: No

17a: $S \cup \{\mathbf{v}\} = \left\{ \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \begin{bmatrix} -4 \\ 4 \end{bmatrix} \right\}$ and $-1 \begin{bmatrix} 2 \\ 2 \end{bmatrix} + 2 \begin{bmatrix} -1 \\ 3 \end{bmatrix} - 1 \begin{bmatrix} -4 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

17d: $S \cup \{\mathbf{v}\} = \left\{1, t, t^2, 5t^2 - 9t + 5\right\}$ and $-5(1) + 9(t) - 5(t^2) + 1(5t^2 - 9t + 5) = 0$

21: line 1: $\mathbf{u}_2 + \mathbf{v} = \mathbf{v}$ for any vector $\mathbf{v}$ (including $\mathbf{u}_1$); line 2: commutativity of addition (vector space definition property 2); line 3: $\mathbf{u}_1 + \mathbf{v} = \mathbf{v}$ for any vector $\mathbf{v}$ (including $\mathbf{u}_2$)

Section 4.2

1b: 5

1f: 6

2c: it is not linearly independent

3e: it is not a subset of $\mathbb{P}_3(\mathbb{R})$.

4b: (i) $\mathbb{P}_2(\mathbb{R})$ (ii) one answer is

$$\left\{9 - 4t + 7t^2, -20 + 8t + 17t^2\right\}$$

(but answers will vary since any two vectors from the set will do)

5b: $\left\{-20 + 8t + 17t^2, -11 + 4t + 24t^2\right\}$ (but answers will vary since any two vectors from the set will do, and it has to be different from exercise 4b); yes, this subset is a basis for the span

6e: 2

6f: 2

8: $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$ must be linearly independent. If it were linearly dependent, we would be able to eliminate one or more of the vectors without affecting the span, resulting in a linearly independent spanning set (basis) with fewer than four elements.

9: $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$ must be spanning. If it were not, there would be a vector, call it $\mathbf{b}_5$, that could not be written as a linear combination of $\mathbf{b}_1$ through $\mathbf{b}_4$, resulting in a linearly independent set, $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4, \mathbf{b}_5\}$, in $\mathbb{R}^4$, contradicting theorem 9.

12d: yes—the determinant of the given matrix is nonzero

12f: no—there are more than four columns

13d: yes–the matrix can be reduced to the (6×6) identity matrix

Section 4.3

1b: yes

1d: no

1f: yes

2b: yes

2d: no

2f: yes

11a: $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$

11b: $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$

12d: $\begin{bmatrix} 114 \\ -65 \\ -104 \end{bmatrix}$

13b: Answers will vary. Any counterexample to (violation of) the two properties of a linear transformation will do. For example, $f(1+1) = \ln(1+1) = \ln 2$ but $f(1)+f(1) = \ln 1+\ln 1 = 0$. f does not preserve addition.

14d: We must demonstrate that the two properties of linear transformations hold.

1.

$$\begin{aligned} &T\left(\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} 4 \\ -1 \\ -5 \end{bmatrix}^T \left(\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} 4 \\ -1 \\ -5 \end{bmatrix}^T \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} 4 \\ -1 \\ -5 \end{bmatrix}^T \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} \\ &= T\left(\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}\right) + T\left(\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}\right) \end{aligned}$$

2.

$$\begin{aligned} T\left(c\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) &= \begin{bmatrix} 4 \\ -1 \\ -5 \end{bmatrix}^T \left(c\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) \\ &= c\begin{bmatrix} 4 \\ -1 \\ -5 \end{bmatrix}^T \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\ &= cT\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) \end{aligned}$$

15b: We must demonstrate that the two properties of linear transformations hold.

1.

$$\begin{aligned} &L\Big((a_1x^2 + b_1x + c_1) + (a_2x^2 + b_2x + c_2)\Big) \\ &= L\Big((a_1 + a_2)x^2 + (b_1 + b_2)x + (c_1 + c_2)\Big) \\ &= 6(a_1 + a_2)x + 3(b_1 + b_2) \\ &= 6a_1x + 6a_2x + 3b_1 + 3b_2 \\ &= (6a_1x + 3b_1) + (6a_2x + 3b_2) \\ &= L(a_1x^2 + b_1x + c_1) + L(a_2x^2 + b_2x + c_2) \end{aligned}$$

2.

$$\begin{aligned} &L\Big(k(ax^2 + bx + c)\Big) \\ &= L(kax^2 + kbx + kc) \\ &= 6kax + 3kb \\ &= k(6ax + 3b) \\ &= kL(ax^2 + bx + c) \end{aligned}$$

18b: $T\left(\begin{bmatrix} 6 \\ -3 \end{bmatrix}\right) = T\left(3\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = 3T\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} -9 \\ -15 \\ 3 \end{bmatrix}$

18d: $T\left(\begin{bmatrix} -3 \\ 5 \end{bmatrix}\right) = T\left(\begin{bmatrix} -1 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = T\left(\begin{bmatrix} -1 \\ 4 \end{bmatrix}\right) - T\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 10 \\ 3 \end{bmatrix}$

19a: $T\left(\begin{bmatrix} 4 \\ 0 \\ -8 \end{bmatrix}\right) = T\left(-2\begin{bmatrix} -2 \\ 0 \\ 4 \end{bmatrix}\right) = -2T\left(\begin{bmatrix} -2 \\ 0 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$

19c: $T\left(\begin{bmatrix}1\\-4\\7\end{bmatrix}\right) = T\left(\begin{bmatrix}5\\-4\\-1\end{bmatrix}+2\begin{bmatrix}-2\\0\\4\end{bmatrix}\right) = T\left(\begin{bmatrix}5\\-4\\-1\end{bmatrix}\right) + T\left(2\begin{bmatrix}-2\\0\\4\end{bmatrix}\right) = T\left(\begin{bmatrix}5\\-4\\-1\end{bmatrix}\right) + 2T\left(\begin{bmatrix}-2\\0\\4\end{bmatrix}\right) = \begin{bmatrix}-5\\1\end{bmatrix}+2\begin{bmatrix}-3\\-5\end{bmatrix} = \begin{bmatrix}-11\\-9\end{bmatrix}$

20: $M = \begin{bmatrix}-2 & -1\\ -\frac{15}{7} & \frac{5}{7}\\ \frac{8}{7} & \frac{9}{7}\end{bmatrix}$

Section 4.4

1a: $\begin{bmatrix}1 & 0\\ -\frac{1}{2} & 1\end{bmatrix}$

1c: $\begin{bmatrix}3 & 0\\ 0 & 1\end{bmatrix}$

3d: $\begin{bmatrix}\frac{28}{9} & \frac{14}{3}\\ 0 & \frac{28}{9}\end{bmatrix}$

3f: $\begin{bmatrix}1 & 0\\ 2 & 0\end{bmatrix}$

3g: $\begin{bmatrix}\frac{17}{6} & \frac{12.92}{6}\\ 0 & 1\end{bmatrix}$

4d: yes

5d: $\begin{bmatrix}-79 & 132\\ 117 & 98\\ 38 & -77\end{bmatrix}$

6b: $\begin{bmatrix}1 & 1.62\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & \frac{11}{7}\end{bmatrix}$

6d: $\begin{bmatrix}1 & 1.27\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$

6f: $\begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & \frac{2}{3}\end{bmatrix}\begin{bmatrix}\frac{2}{3} & 0\\ 0 & 1\end{bmatrix}$

7c: $\begin{bmatrix}42 & -641 & -135\\ -1 & 8 & 2\\ -12 & 258 & 52\\ -24 & 282 & 62\end{bmatrix}$

7e: $\begin{bmatrix}-172 & 23 & 33 & 151\\ -1 & 2 & -2 & 15\end{bmatrix}$

10: $M = \begin{bmatrix}-1 & 0\\ 0 & -1\end{bmatrix}$

13c: $\begin{bmatrix}-1 & 0\\ 0 & -1\end{bmatrix}$

14b: $\begin{bmatrix}0 & 0\\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}\end{bmatrix}$

Section 4.5

10e: (i) Yes. (ii) Yes. (iii) Yes.

12b: $h\left(\begin{bmatrix}r_1 & r_2 & \cdots & r_n\end{bmatrix}\right) = \langle r_1, r_2, \ldots, r_n\rangle$

12d: $p\left(r_1 I_{:,1} + r_2 I_{:,2} + \cdots + r_n I_{:,n}\right) = (r_1, r_2, \ldots, r_n)$

13a: By definition, $T_M : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one precisely when

$$M\mathbf{v} = \mathbf{b}$$

has at most one solution for each $\mathbf{b}$ in $\mathbb{R}^m$ (they are equivalent). Since the latter is one of the statements of the theorem, "$T_M : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one" may be added as yet another equivalent statement.

13c: By definition, $T_M : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one precisely when

$$M\mathbf{v} = \mathbf{b}$$

has at most one solution for each $\mathbf{b}$ in $\mathbb{R}^m$ (they are equivalent). By definition, $T_M : \mathbb{R}^n \to \mathbb{R}^m$ is onto precisely when

$$M\mathbf{v} = \mathbf{b}$$

has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^m$ (they are equivalent). Since the latter halves of each of these statements appear in the theorem, the properties "$T_M : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one" and "$T_M : \mathbb{R}^n \to \mathbb{R}^m$ is onto" may be added as additional equivalent statements.

13e: Let $f \subseteq V \times W$ be an isomorphism. By definition, f is one-to-one and onto. Therefore the equation $f(a) = b$ has at most one solution for each b (by definition of one-to-one) and has at least one solution for each b (by definition of onto). Since the number of solutions is both less than or equal to one and greater than or equal to one, it must be exactly one (for each b). Hence each element b of W has exactly one preimage a in V. In other words, the relation $f^{-1} \subseteq W \times V$ contains exactly one element (b, a) for each b in W. Therefore the domain of f^{-1} is W and f^{-1} is a one-to-one function. That f^{-1} is onto follows from the fact that f is a function (there is a pair (a, b) in f for every element a of the domain V so there is a pair (b, a) in f^{-1} for every element a of V). Since f^{-1} is a one-to-one and onto function, it is an isomorphism.

Section 4.6

2c: 135

2f: 113

3a: 1934

3d: $\sqrt{34}$

4b: $\frac{1}{\pi}\left(\pi + 20\frac{\ln(10-\pi)}{\ln 10}\right) = 1 + \frac{20}{\pi}\log(10-\pi)$

4d: $\frac{\sqrt{2}}{2}$

5a: -50

5d: $\sqrt{140}$

6d: $\sqrt{85}$

7d: $\sqrt{290}$

8a: $\frac{1}{2}$

9d: $2\sqrt{142}$

14a: All scalar multiples of $\begin{bmatrix} 5 \\ -2 \end{bmatrix}$

14c: All functions whose integral over $[0, 1]$ is zero. In other words any function whose graph has an area above the x-axis equal to the area below the x-axis over the interval $[0, 1]$.

16b: $\frac{11}{3}$

16d: 2

18: HINT: It fails to satisfy inner product property 2. Can you show this?

Section 5.1

1e: yes

1f: no

2e: Answers may vary. One solution is $\left\{\begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} -6 \\ -14 \end{bmatrix}\right\}$.

2f: Answers may vary. One solution is $\left\{\begin{bmatrix} 10 \\ -40 \end{bmatrix}\right\}$.

3e: {} (the empty set, whose span is $\{\mathbf{0}\}$)

3f: Answers may vary. One solution is $\left\{\begin{bmatrix} -3 \\ 5 \end{bmatrix}\right\}$.

4f: (i) yes (ii) Answers may vary. One solution is $\left\{\begin{bmatrix} -12 \\ -36 \\ 24 \\ 60 \end{bmatrix}, \begin{bmatrix} 6 \\ 15 \\ -9 \\ -30 \end{bmatrix}, \begin{bmatrix} 11 \\ 33 \\ -11 \\ -33 \end{bmatrix}\right\}$ (iii) Answers may vary. One solution is $\left\{\begin{bmatrix} 3 \\ -16 \\ 6 \\ 0 \end{bmatrix}\right\}$

4h: (i) yes (ii) Answers may vary. One solution is $\left\{\begin{bmatrix} -12 \\ -36 \\ 24 \\ -24 \\ 36 \end{bmatrix}, \begin{bmatrix} -8 \\ -20 \\ 24 \\ -8 \\ 32 \end{bmatrix}, \begin{bmatrix} 8 \\ 16 \\ 24 \\ -16 \\ -72 \end{bmatrix}, \begin{bmatrix} -48 \\ -132 \\ 156 \\ -84 \\ 132 \end{bmatrix}\right\}$ (iii) {} (the empty set, whose span is $\{\mathbf{0}\}$)

4j: (i) yes (ii) Any five linearly independent vectors in $\mathbb{R}^5$ will do. For example, the columns of M or the standard basis. (iii) {} (the empty set, whose span is $\{\mathbf{0}\}$)

6d: (i) $\mathbf{b} = -\frac{1}{7}M_{:,2}$ (ii) $\mathbf{v} = \begin{bmatrix} 0 \\ -\frac{1}{7} \\ 0 \\ 0 \end{bmatrix} + r\begin{bmatrix} 0 \\ -4 \\ 7 \\ 0 \end{bmatrix}$

13: Columns 1,3, and 5 are linearly independent. Yes, there are other such sets.

15d: 1

18: 1

Section 5.2

5b: $\mathbf{v} = \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix}_{\mathcal{B}}$

6d: $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}_{\mathcal{B}}$

7f: $\mathbf{v} = [12]_{\mathcal{B}}$

9: (a) $\mathbf{v} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}_{\mathcal{B}}$ (b) $\mathbf{v} = \begin{bmatrix} -3 \\ -2 \end{bmatrix}_{\mathcal{B}}$

11: (a) $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}_{\mathcal{B}}$ (b) $\mathbf{v} = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix}_{\mathcal{B}}$

(c) $\mathbf{v} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}_{\mathcal{B}}$

15: $[\mathcal{B}]_{\mathcal{E}} = \begin{bmatrix} 6 & -5 \\ 3 & -8 \end{bmatrix}$ (a) $[\mathcal{B}]_{\mathcal{E}}^{-1}\begin{bmatrix} 17 \\ 14 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ (b) $[\mathcal{B}]_{\mathcal{E}}^{-1}\begin{bmatrix} -8 \\ 7 \end{bmatrix} = \begin{bmatrix} -3 \\ -2 \end{bmatrix}_{\mathcal{B}}$; The coordinates are the same as those in question 9.

17: $[\mathcal{B}]_{\mathcal{E}} = \begin{bmatrix} -4 & 9 & -3 \\ 7 & 4 & 5 \\ 3 & 3 & 4 \end{bmatrix}$

(a) $[\mathcal{B}]_{\mathcal{E}}^{-1}\begin{bmatrix} 17 \\ 10 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$

(b) $[\mathcal{B}]_{\mathcal{E}}^{-1}\begin{bmatrix} 8 \\ 5 \\ 9 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix}_{\mathcal{B}}$

(c) $[\mathcal{B}]_{\mathcal{E}}^{-1}\begin{bmatrix} 19 \\ 6 \\ 7 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}_{\mathcal{B}}$; The coordinates are the same as those in question 11.

21: $[\mathcal{B}]_{C} = \begin{bmatrix} -\frac{4}{5} & \frac{3}{2} \\ \frac{11}{15} & 1 \end{bmatrix}$

23: $[\mathcal{B}]_{C} = \begin{bmatrix} -\frac{1}{2} & -\frac{155}{136} & \frac{7}{68} \\ \frac{1}{2} & -\frac{43}{136} & -\frac{45}{68} \\ 0 & -\frac{38}{17} & -\frac{23}{17} \end{bmatrix}$ as produced by the code at SageMathCell 139.

27: (a) $\mathbf{v} = \begin{bmatrix} \frac{3}{19} \\ \frac{58}{57} \end{bmatrix}_{\mathcal{B}}$ (b) $\mathbf{v} = \begin{bmatrix} \frac{7}{5} \\ \frac{17}{15} \end{bmatrix}_{C}$

(c) $\begin{bmatrix} -\frac{4}{5} & \frac{3}{2} \\ \frac{11}{15} & 1 \end{bmatrix}\begin{bmatrix} \frac{3}{19} \\ \frac{58}{57} \end{bmatrix} = \begin{bmatrix} \frac{7}{5} \\ \frac{17}{15} \end{bmatrix}$ which is $[\mathbf{v}]_C$ as expected

29: As produced by the code at SageMathCell 140,

(a) $\mathbf{v} = \begin{bmatrix} \frac{55}{49} \\ -\frac{9}{49} \\ \frac{17}{49} \end{bmatrix}_{\mathcal{B}}$ (b) $\mathbf{v} = \begin{bmatrix} -\frac{43}{136} \\ \frac{53}{136} \\ -\frac{1}{17} \end{bmatrix}_{C}$ (c) $\begin{bmatrix} -\frac{1}{2} & -\frac{155}{136} & \frac{7}{68} \\ \frac{1}{2} & -\frac{43}{136} & -\frac{45}{68} \\ 0 & -\frac{38}{17} & -\frac{23}{17} \end{bmatrix}\begin{bmatrix} \frac{55}{49} \\ -\frac{9}{49} \\ \frac{17}{49} \end{bmatrix} = \begin{bmatrix} -\frac{43}{136} \\ \frac{53}{136} \\ -\frac{1}{17} \end{bmatrix}$ which is $[\mathbf{v}]_C$ as expected

Section 5.3

1e: yes

1f: no. $\begin{bmatrix} \frac{1}{\sqrt{35}} & \frac{3}{\sqrt{35}} & -\frac{3}{\sqrt{35}} & \frac{4}{\sqrt{35}} \end{bmatrix}$

2e: S is orthogonal

2i: S is not orthogonal (Orthogonal sets of nonzero vectors are linearly independent, but this set is necessarily linearly dependent having four vectors from $\mathbb{R}^3$.)

3e: S is not orthogonal

3g: S is not orthogonal

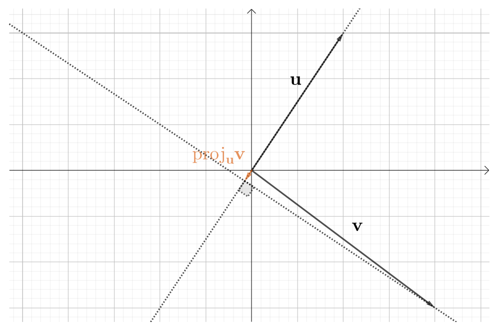
3i: S is not orthogonal (Orthogonal sets of nonzero vectors are linearly independent, but this set is necessarily linearly dependent having four vectors from $\mathbb{R}^3$.)

5d: $4t^2 - 8t + 3$ is not orthogonal to $t^2 - 2t$.

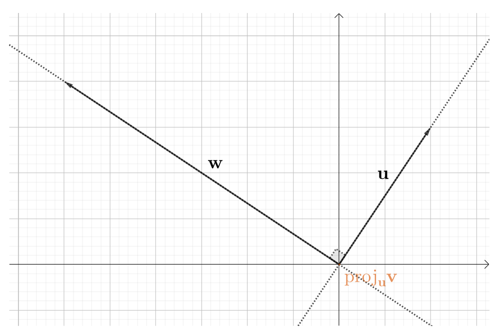
6c: $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$

6e: $\begin{bmatrix} -\frac{4}{25} \\ \frac{3}{25} \end{bmatrix}$

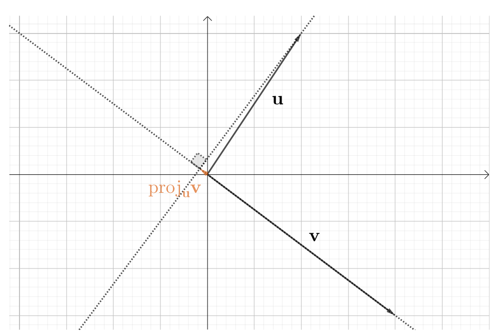
7a:



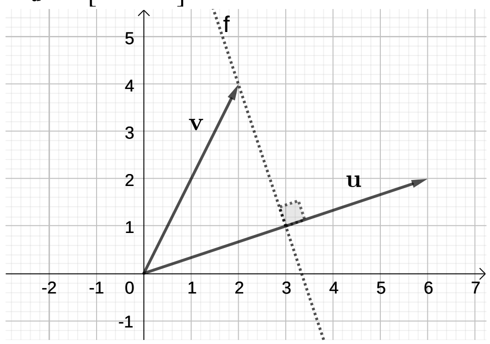
7c:



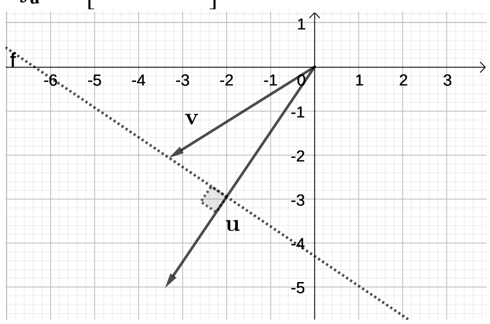
7e:



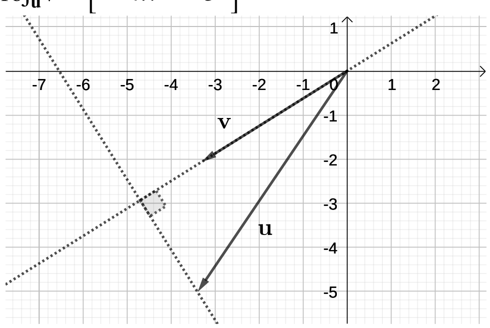
8a: $\text{proj}_{\mathbf{u}}\mathbf{v} = \begin{bmatrix} 3 & 1 \end{bmatrix}^T$



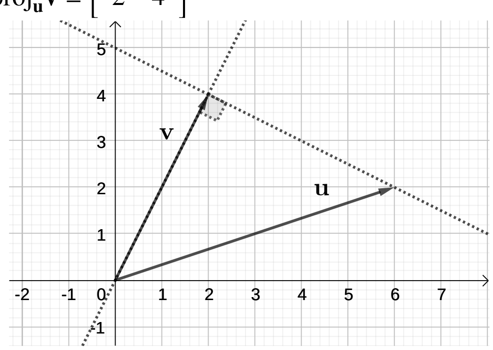
8c: $\text{proj}_{\mathbf{u}}\mathbf{v} = \begin{bmatrix} -2 & -3 \end{bmatrix}^T$



9a: $\text{proj}_{\mathbf{u}}\mathbf{v} = \begin{bmatrix} -4.7 & -3 \end{bmatrix}^T$



9c: $\text{proj}_{\mathbf{u}}\mathbf{v} = \begin{bmatrix} 2 & 4 \end{bmatrix}^T$



11b: $\mathbf{v} = \begin{bmatrix} -\frac{61}{29} \\ \frac{22}{29} \end{bmatrix}$

12c: Answers may vary. One answer is $\left\{ \begin{bmatrix} 8 \\ -9 \end{bmatrix}, \begin{bmatrix} 9 \\ 8 \end{bmatrix} \right\}$. Multiples of these vectors also suffice, for example. Actually, *any* orthogonal set of two vectors, such as the standard basis, will do since the span of the given set is $\mathbb{R}^2$.

12e: Answers may vary. One answer is $\left\{ \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 3 \end{bmatrix} \right\}$. Multiples of these vectors also suffice, for example.

12g: Answers may vary. One answer is $\left\{ \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 9 \\ -6 \\ -3 \end{bmatrix} \right\}$. Multiples of these vectors also suffice, for example. Actually, *any* orthogonal set of two vectors, such as the standard basis, will do since the span of the given set is $\mathbb{R}^3$.

12i: Answers may vary. One answer is $\left\{ \begin{bmatrix} -2 \\ -4 \\ 7 \\ 3 \end{bmatrix}, \begin{bmatrix} 35 \\ 109 \\ 53 \\ 45 \end{bmatrix}, \begin{bmatrix} 685 \\ -377 \\ -369 \\ 815 \end{bmatrix}, \begin{bmatrix} 29174 \\ -7042 \\ 13581 \\ -21629 \end{bmatrix} \right\}$ as shown at SageMathCell 141. Multiples of these vectors also suffice, for example. Actually, *any* orthogonal set of two vectors, such as the standard basis, will do since the span of the given set is $\mathbb{R}^4$.

13c: As shown at SageMathCell 142,

$$\left\{ \begin{bmatrix} \frac{2}{\sqrt{21}} \\ \frac{1}{\sqrt{21}} \\ \frac{4}{\sqrt{21}} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{3}{\sqrt{14}} \\ -\frac{2}{\sqrt{14}} \\ -\frac{1}{\sqrt{14}} \end{bmatrix} \right\}$$

14a: yes; $\mathbb{R}^2$

14c: yes; $\mathbb{R}^2$

14e: no

14g: yes; $\mathbb{R}^3$

14i: yes; $\mathbb{R}^4$

15a: $\left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \right\}$

15c: $\left\{ \begin{bmatrix} \frac{8}{\sqrt{145}} \\ -\frac{9}{\sqrt{145}} \end{bmatrix}, \begin{bmatrix} \frac{9}{\sqrt{145}} \\ \frac{8}{\sqrt{145}} \end{bmatrix} \right\}$

15e: $\left\{ \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}$

15g: $\left\{ \begin{bmatrix} \frac{2}{\sqrt{21}} \\ \frac{1}{\sqrt{21}} \\ \frac{4}{\sqrt{21}} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} \frac{3}{\sqrt{14}} \\ -\frac{2}{\sqrt{14}} \\ -\frac{1}{\sqrt{14}} \end{bmatrix} \right\}$

15i: $\left\{ \begin{bmatrix} -\frac{2}{\sqrt{78}} \\ -\frac{4}{\sqrt{78}} \\ \frac{7}{\sqrt{78}} \\ \frac{3}{\sqrt{78}} \end{bmatrix}, \begin{bmatrix} \frac{35}{2\sqrt{4485}} \\ \frac{109}{2\sqrt{4485}} \\ \frac{53}{2\sqrt{4485}} \\ \frac{45}{2\sqrt{4485}} \end{bmatrix}, \begin{bmatrix} \frac{685}{6\sqrt{39215}} \\ -\frac{377}{6\sqrt{39215}} \\ -\frac{123}{2\sqrt{39215}} \\ \frac{815}{6\sqrt{39215}} \end{bmatrix}, \begin{bmatrix} \frac{58}{3\sqrt{682}} \\ -\frac{14}{3\sqrt{682}} \\ \frac{9}{\sqrt{682}} \\ -\frac{43}{3\sqrt{682}} \end{bmatrix} \right\}$

16e: S is orthogonal as-is. Normalized:

$$\left\{ \begin{bmatrix} \frac{6}{\sqrt{46}} \\ -\frac{1}{\sqrt{46}} \\ \frac{3}{\sqrt{46}} \end{bmatrix}, \begin{bmatrix} -\frac{4}{\sqrt{122}} \\ -\frac{9}{\sqrt{122}} \\ \frac{5}{\sqrt{122}} \end{bmatrix} \right\}$$

16g: S is orthogonal as-is. Normalized:

$$\left\{ \begin{bmatrix} \frac{2}{\sqrt{29}} \\ \frac{3}{\sqrt{29}} \\ \frac{4}{\sqrt{29}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{165}} \\ -\frac{10}{\sqrt{165}} \\ \frac{8}{\sqrt{165}} \end{bmatrix}, \begin{bmatrix} \frac{64}{\sqrt{4785}} \\ -\frac{20}{\sqrt{4785}} \\ -\frac{17}{\sqrt{4785}} \end{bmatrix} \right\}$$

[Technically, any orthogonal set of at least 3 vectors will suffice since any such set spans $\mathbb{R}^3$. We have not been asked to use the orthogonalization procedure.]

16i: Following the orthogonalization process leads to

$$\left\{ \begin{bmatrix} -7 \\ 0 \\ 9 \end{bmatrix}, \begin{bmatrix} 9 \\ 8 \\ 7 \end{bmatrix}, \begin{bmatrix} 288 \\ -520 \\ 224 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

[Technically, any orthogonal set of at least 3 vectors will suffice since any such set spans $\mathbb{R}^3$. We have not been asked to use the orthogonalization procedure.] The zero vector cannot be normalized so must not be included in the orthonormal set. Normalizing the other three vectors:

$$\left\{ \begin{bmatrix} -\frac{7}{\sqrt{130}} \\ 0 \\ \frac{9}{\sqrt{130}} \end{bmatrix}, \begin{bmatrix} \frac{9}{\sqrt{194}} \\ \frac{8}{\sqrt{194}} \\ \frac{7}{\sqrt{194}} \end{bmatrix}, \begin{bmatrix} \frac{36}{\sqrt{6305}} \\ -\frac{65}{\sqrt{6305}} \\ \frac{28}{\sqrt{6305}} \end{bmatrix} \right\}.$$

[Technically, any orthonormal set of 3 vectors, such as the standard basis, will suffice since any such set spans $\mathbb{R}^3$. We have not been asked to use the orthogonalization procedure.]

17e: Orthogonalized: $\left\{ \begin{bmatrix} 6 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} -19 \\ -21 \\ 8 \end{bmatrix} \right\}$;

Orthonormalized: $\left\{ \begin{bmatrix} \frac{6}{\sqrt{65}} \\ -\frac{1}{\sqrt{65}} \\ \frac{3}{\sqrt{65}} \end{bmatrix}, \begin{bmatrix} -\frac{19}{\sqrt{1435}} \\ -\frac{21}{\sqrt{1435}} \\ \frac{8}{\sqrt{1435}} \end{bmatrix} \right\}$

17g: Orthogonalized: $\left\{ \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -69 \\ -401 \\ 212 \end{bmatrix}, \begin{bmatrix} 192 \\ -30 \\ -17 \end{bmatrix} \right\}$;

Orthonormalized:

$$\left\{ \begin{bmatrix} \frac{2}{\sqrt{70}} \\ \frac{3}{\sqrt{70}} \\ \frac{4}{\sqrt{70}} \end{bmatrix}, \begin{bmatrix} -\frac{69}{\sqrt{461195}} \\ -\frac{401}{\sqrt{461195}} \\ \frac{212}{\sqrt{461195}} \end{bmatrix}, \begin{bmatrix} \frac{192}{\sqrt{39531}} \\ -\frac{30}{\sqrt{39531}} \\ -\frac{17}{\sqrt{39531}} \end{bmatrix} \right\}$$

17i: Orthogonalized:

$$\left\{ \begin{bmatrix} -7 \\ 0 \\ 9 \end{bmatrix}, \begin{bmatrix} 1755 \\ 1168 \\ 455 \end{bmatrix}, \begin{bmatrix} 216 \\ -195 \\ 56 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\};$$

Orthonormalized:

$$\left\{ \begin{bmatrix} -\frac{7}{2\sqrt{73}} \\ 0 \\ \frac{9}{2\sqrt{73}} \end{bmatrix}, \begin{bmatrix} \frac{1755}{2\sqrt{1607387}} \\ \frac{584}{\sqrt{1607387}} \\ \frac{455}{2\sqrt{1607387}} \end{bmatrix}, \begin{bmatrix} \frac{216}{\sqrt{132114}} \\ -\frac{195}{\sqrt{132114}} \\ \frac{56}{\sqrt{132114}} \end{bmatrix} \right\}$$

20: $\left\{1,\ t-1,\ t^2-2t+\frac{1}{3}\right\}$

Section 5.4

1c: Answers may vary. $P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$; $P^{-1}MP = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

1e: Answers may vary. $P = \begin{bmatrix} 0 & 4 & 0 \\ -1 & 1 & -1 \\ 2 & 3 & 1 \end{bmatrix}$; $P^{-1}MP = \begin{bmatrix} -8 & 0 & 0 \\ 0 & -20 & 0 \\ 0 & 0 & -16 \end{bmatrix}$

1i: Answers may vary. $P = \begin{bmatrix} 0 & 1 & 3 & 2 \\ 3 & 3 & 1 & -3 \\ -2 & -2 & -2 & 0 \\ 1 & -2 & -1 & 1 \end{bmatrix}$;

$$P^{-1}MP = \begin{bmatrix} 7 & 0 & 0 & 0 \\ 0 & 21 & 0 & 0 \\ 0 & 0 & -7 & 0 \\ 0 & 0 & 0 & -7 \end{bmatrix}$$

2c: $-1, 1$

2e: $-20, -16, -8$

2i: $-7, 7, 21$

3a: no

3h: no

4d: $k = 0$

5d: Answers will vary. $P = \begin{bmatrix} 1 & 4 \\ 1 & 5 \end{bmatrix}$

10c: Answers will vary. $P = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ where

$$a = -\frac{25698}{97469}r_1 + \frac{319807}{97469}r_2 - \frac{50820}{97469}r_3$$

$$b = -\frac{16879}{389876}r_1 + \frac{1996453}{389876}r_2 - \frac{317321}{389876}r_3$$

$$c = r_1$$

$$d = \frac{326995}{389876}r_1 - \frac{319057801}{389876}r_2 + \frac{50321237}{389876}r_3$$

$$e = \frac{13066}{97469}r_1 - \frac{12604743}{97469}r_2 + \frac{1988259}{97469}r_3$$

$$f = -\frac{1244091}{389876}r_1 + \frac{1215417989}{389876}r_2 - \frac{191691365}{389876}r_3$$

$$g = r_3$$

$$h = r_2$$

$$i = -\frac{41}{389876}r_1 + \frac{1106915}{389876}r_2 - \frac{1659967}{389876}r_3$$

11a: 2

11c: 3

12b: Reflection across the line $y = \frac{1}{2}x$ parallel to the y-axis.

12c: Scaling by a factor of 2 in the direction of $\begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}$ and by a factor of 3 in the direction of $\begin{bmatrix} -\sin\alpha \\ \cos\alpha \end{bmatrix}$.

13d: $M^7 = \begin{bmatrix} 1 & 1 \\ -7 & 1 \end{bmatrix} \begin{bmatrix} \frac{6}{5} & 0 \\ 0 & \frac{4}{5} \end{bmatrix}^7 \frac{1}{8} \begin{bmatrix} 1 & -1 \\ 7 & 1 \end{bmatrix}$

$= \frac{1}{8} \begin{bmatrix} \left(\frac{6}{5}\right)^7 + 7\left(\frac{4}{5}\right)^7 & -\left(\frac{6}{5}\right)^7 + \left(\frac{4}{5}\right)^7 \\ -7\left(\frac{6}{5}\right)^7 + 7\left(\frac{4}{5}\right)^7 & 7\left(\frac{6}{5}\right)^7 + \left(\frac{4}{5}\right)^7 \end{bmatrix}$

Section 6.1

1a: Answers may vary. One solution is $L = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$; $U = \begin{bmatrix} 3 & 5 \\ 0 & 1 \end{bmatrix}$

1c: Answers may vary. One solution is $L = \begin{bmatrix} 1 & 0 \\ -\frac{5}{2} & \frac{1}{2} \end{bmatrix}$; $U = \begin{bmatrix} -2 & -6 \\ 0 & 40 \end{bmatrix}$

1g: Answers may vary. One solution is $L = \begin{bmatrix} -5 & 0 & 0 \\ -4 & -21 & 0 \\ 1 & -6 & 1 \end{bmatrix}$; $U = \begin{bmatrix} 5 & -2 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

1i: Answers may vary. One solution is $L = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ \frac{3}{2} & -\frac{3}{2} & \frac{1}{2} \end{bmatrix}$; $U = \begin{bmatrix} -6 & 8 & 4 \\ 0 & 12 & 0 \\ 0 & 0 & -12 \end{bmatrix}$

2a: Answers may vary. One solution is $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$; $L = \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix}$; $U = \begin{bmatrix} 1 & 7 \\ 0 & 25 \end{bmatrix}$

2e: Answers may vary. One solution is $P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$; $L = \begin{bmatrix} 1 & 0 & 0 \\ 8 & 1 & 0 \\ 12 & -2 & 1 \end{bmatrix}$; $U = \begin{bmatrix} -1 & 2 & 0 \\ 0 & 6 & -4 \\ 0 & 0 & -7 \end{bmatrix}$

3a: $\det M = 3 \cdot 1 \cdot 1 \cdot 1 = 3$

3c: $\det M = 1 \cdot \frac{1}{2} \cdot -2 \cdot 40 = -40$

3g: M is not square so has no determinant

3i: $\det M = 1 \cdot 1 \cdot \frac{1}{2} \cdot -6 \cdot 12 \cdot -12 = 432$

4a: $\det M = -1 \cdot 1 \cdot 1 \cdot 1 \cdot 25 = -25$

4e: $\det M = -1 \cdot 1 \cdot 1 \cdot 1 \cdot -1 \cdot 6 \cdot -7 = -42$

5a: $L = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$; $U = \begin{bmatrix} 3 & 5 \\ 0 & 1 \end{bmatrix}$

5c: $L = \begin{bmatrix} 1 & 0 \\ -\frac{5}{2} & 1 \end{bmatrix}$; $U = \begin{bmatrix} -2 & -6 \\ 0 & 20 \end{bmatrix}$

5g: $L = \begin{bmatrix} 1 & 0 & 0 \\ \frac{4}{5} & 1 & 0 \\ -\frac{1}{5} & \frac{2}{7} & 1 \end{bmatrix}$; $U = \begin{bmatrix} -25 & 10 \\ 0 & -21 \\ 0 & 0 \end{bmatrix}$

5i: $L = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ \frac{3}{2} & -\frac{3}{2} & 1 \end{bmatrix}$; $U = \begin{bmatrix} -6 & 8 & 4 \\ 0 & 12 & 0 \\ 0 & 0 & -6 \end{bmatrix}$

6a: $L = \begin{bmatrix} 3 & 0 \\ -9 & 1 \end{bmatrix}$; $U = \begin{bmatrix} 1 & \frac{5}{3} \\ 0 & 1 \end{bmatrix}$

6c: $L = \begin{bmatrix} -2 & 0 \\ 5 & 20 \end{bmatrix}$; $U = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$

6e: $\begin{bmatrix} 4 & 0 \\ 1 & -36 \end{bmatrix} \begin{bmatrix} 1 & 6 & 6 \\ 0 & 1 & \frac{1}{6} \end{bmatrix}$

6g: $L = \begin{bmatrix} -25 & 0 & 0 \\ -20 & -21 & 0 \\ 5 & -6 & 1 \end{bmatrix}$; $U = \begin{bmatrix} 1 & -\frac{2}{5} \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

6i: $L = \begin{bmatrix} -6 & 0 & 0 \\ 6 & 12 & 0 \\ -9 & -18 & -6 \end{bmatrix}$; $U = \begin{bmatrix} 1 & -\frac{8}{6} & -\frac{4}{6} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

7b: $\mathbf{v} = \begin{bmatrix} 2 \\ -13 \end{bmatrix}$

7f: $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$

9b: $M^{-1} = \begin{bmatrix} -\frac{1}{15} & -\frac{1}{30} \\ \frac{2}{3} & -\frac{1}{6} \end{bmatrix}$

9d: $M^{-1} = \begin{bmatrix} -\frac{4}{3} & -\frac{2}{3} & \frac{1}{2} \\ \frac{1}{7} & 0 & \frac{1}{14} \\ -3 & -1 & \frac{1}{2} \end{bmatrix}$

9f: $M^{-1} = \begin{bmatrix} -\frac{33}{2} & -\frac{15}{2} & 1 \\ -\frac{2}{7} & -\frac{1}{7} & 0 \\ -8 & -\frac{7}{2} & \frac{1}{2} \end{bmatrix}$

Section 6.2

1d: 7

1f: −6

1h: −24

2d: no

2f: no

3c: Answers may vary. One answer is 3.996, $\begin{bmatrix} 16381 \\ 8191 \end{bmatrix}$

3e: Answers may vary. One answer is 83.06, $\begin{bmatrix} 10654688 \\ 16132160 \\ -10084800 \end{bmatrix}$

4c: $\mathbf{v}_8$ through $\mathbf{v}_{11}$ are

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 17 \\ -8 \end{bmatrix}, \begin{bmatrix} 49 \\ 0 \end{bmatrix}, \begin{bmatrix} 833 \\ -392 \end{bmatrix}$$

giving two rather different directions, $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 2.125 \\ -1 \end{bmatrix}$, so it seems the method will not converge.

4e: $\mathbf{v}_8$ through $\mathbf{v}_{11}$ are

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -11 \\ -14 \\ -13 \end{bmatrix}, \begin{bmatrix} 81 \\ 81 \\ 81 \end{bmatrix}, \begin{bmatrix} -891 \\ -1134 \\ -1053 \end{bmatrix}$$

giving two rather different directions, $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 11 \\ 14 \\ 13 \end{bmatrix}$, so it seems the method will not converge.

4g: $\mathbf{v}_{10} = \begin{bmatrix} -53291821 \\ 63954728 \\ -13570337 \end{bmatrix}$ and $\mathbf{v}_{11} = \begin{bmatrix} 3533661079 \\ -4085829332 \\ 896471030 \end{bmatrix}$, both pointing in approximately the direction $\begin{bmatrix} 1 \\ -1.2 \\ 0.25 \end{bmatrix}$, so it seems the method will converge.

5c: 4, $\begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}$

5e: 110, $\begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ -1 \end{bmatrix}$

6b: (i) −12 (ii) 2 (iii) yes, $\begin{bmatrix} -.39 & .8 & -.98 & 1 \end{bmatrix}^T$ (iv) Answers may vary. With $\mathbf{v}_0 = \begin{bmatrix} 8 & -3 & 2 & -7 \end{bmatrix}^T$ yes, it produces the same eigenvector as before.

7b: yes, it works

8: (a) 390 is the dominant eigenvalue of M (b) the eigenvector corresponding with eigenvalue 390 is $\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T$ (c) the power method will work

10a: eigenvalues 33 and 11. It seemed the method would converge, and this is because M has a dominant eigenvalue.

10c: eigenvalues 7 and −7. It seemed the method would not converge, and this is because M does not have a dominant eigenvalue.

10e: eigenvalues 9 and −9. It seemed the method would not converge, and this is because M does not have a dominant eigenvalue.

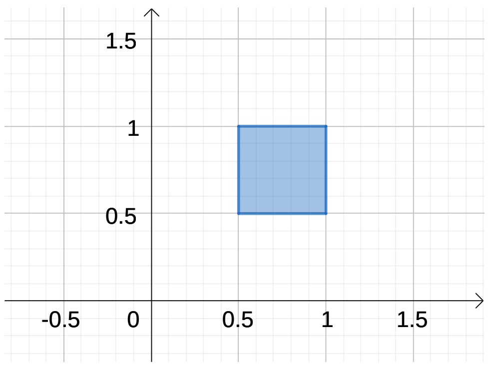
10g: eigenvalues $-39, -52,$ and -65. It seemed the method would converge, and this is because M has a dominant eigenvalue.

Section 6.3

4: 2

5: 1 and 2

9b:



11c: Answers may vary. One solution is $P = \begin{bmatrix} 3 & -8 \\ 0 & 9 \end{bmatrix}$

11f: Answers may vary. One solution is $P = \begin{bmatrix} 2 & 0 & 0 \\ 4 & 6 & 0 \\ -1 & -5 & 1 \end{bmatrix}$

Section 6.4

1d: $\frac{53}{75}\mathbf{v}$

1f: $\frac{102}{269}\mathbf{v}$

1h: $-\frac{47}{315}\mathbf{v}$

2d: 11

2f: $\sqrt{\frac{10819}{130}} \approx 9.123$

2h: $\sqrt{\frac{1745}{201}} \approx 2.946$

3a: $\begin{bmatrix} -\frac{65}{34} \\ \frac{39}{34} \end{bmatrix}$

3c: $\begin{bmatrix} -\frac{106}{29} \\ 0 \\ \frac{265}{29} \end{bmatrix}$

3g: $\begin{bmatrix} \frac{613}{51} \\ \frac{665}{153} \\ -\frac{1147}{153} \\ \frac{838}{153} \end{bmatrix}$

4d: $\begin{bmatrix} -\frac{101217}{19274} \\ -\frac{79991}{19274} \\ \frac{45753}{9637} \\ -\frac{22176}{9637} \end{bmatrix}$

4f: $\begin{bmatrix} 6 \\ 10 \\ 0 \\ 1 \end{bmatrix}$

5d: no

6d:

$$\begin{bmatrix} 7 \\ 7 \\ 12 \end{bmatrix} = \begin{bmatrix} -\frac{17}{194} \\ \frac{54}{97} \\ -\frac{47}{194} \end{bmatrix} + \begin{bmatrix} \frac{1375}{194} \\ \frac{625}{97} \\ \frac{2375}{194} \end{bmatrix}$$

where $\begin{bmatrix} -\frac{17}{194} \\ \frac{54}{97} \\ -\frac{47}{194} \end{bmatrix}$ is in W and $\begin{bmatrix} \frac{1375}{194} \\ \frac{625}{97} \\ \frac{2375}{194} \end{bmatrix}$ is in $W^{\perp}$.

7c: The system is consistent:
$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{1129}\begin{bmatrix} -661 \\ -337 \\ 130 \end{bmatrix}$

7e: The system is consistent:
$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \\ 0 \end{bmatrix} + r\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$

7g: The system is inconsistent. The best approximation of $\mathbf{b} = \begin{bmatrix} 10 \\ 4 \\ 9 \end{bmatrix}$ as a linear combination of the columns of the coefficient matrix is $\begin{bmatrix} \frac{7045}{679} \\ \frac{6605}{1358} \\ \frac{10845}{1358} \end{bmatrix}$. One particular linear combination (best approximation to a solution) is $x = -\frac{935}{679}$, $y = -\frac{285}{194}$, $z = 0$. Answers may vary. There are infinitely many others.

8d: $\left\{\begin{bmatrix} 114 \\ 7 \\ 61 \end{bmatrix}\right\}$

10: See the solution of exercise 3e for a hint.

12: See the solution of exercise 8b for a hint.

15a: $-\frac{187}{274}\left(-6x^2 + 2x - 1\right)$

15c: $-\frac{31}{131}\left(11x^2 + 3x - 1\right)$
$+\frac{503}{979}\left(-4x^3 + 8x^2 + 17x - 28\right)$
$= -\frac{2012}{979}x^3 + \frac{193305}{128249}x^2 + \frac{1029134}{128249}x - \frac{1814655}{128249}$

15e: $-\frac{159}{163}\left(5x^2 - 11x + 14\right) + \frac{124}{121}\left(4x^2 - 12x - 5\right)$
$= -\frac{15347}{19723}x^2 - \frac{30915}{19723}x - \frac{370406}{19723}$

16b: $\frac{73}{50}\left(-3x^2 + 3x - 1\right)$
$-\frac{859}{3382}\left(-14x^3 + 9x^2 - 3x + 25\right)$
$= \frac{6013}{1691}x^3 - \frac{281802}{42275}x^2 + \frac{217377}{42275}x - \frac{330159}{42275}$

17a: no

18a: $\mathbf{w} = \text{proj}_W \mathbf{p} = \frac{520}{857}\left(7x^3 + 8x^2 - 8x - 1\right)$
$= \frac{3640}{857}x^3 + \frac{4160}{857}x^2 - \frac{4160}{857}x - \frac{520}{857}$
$\mathbf{w}^{\perp} = \mathbf{p} - \mathbf{w}^{\perp} = \frac{645}{857}x^3 - \frac{5017}{857}x^2 + \frac{6731}{857}x + \frac{2234}{857}$

19b: $\left\{13 - 6x + 9x^2 - 6x^3,\ -3 + x - 3x^2 + 2x^3\right\}$

Section 7.1

1b: nonlinear

1d: linear

1f: linear

1h: linear

2a: The general shape of the graph is parabolic, so one might try a model of the form $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$

2c: The general shape of the graph is logarithmic, so one might try a model of the form $f(x) = \beta_0 + \beta_1 \ln x$

2e: The general shape of the graph is exponential, so one might try a model of the form $f(x) = \beta_0 + \beta_1\left(\sqrt{2}\right)^x$. See question 6 for further discussion.

3b: $g(x) \approx 2.90418 + 6.06618x$

3f: $\ell(x) \approx \dfrac{50.3321176x}{1 + e^t}$

4b: $g(x) \approx 2.90418 + 6.06618x$

4f: $\ell(x) \approx \dfrac{50.3321176x}{1 + e^t}$

5b: $||M\hat{\mathbf{v}} - \mathbf{b}||^2 \approx 114.654$

5f: $||M\hat{\mathbf{v}} - \mathbf{b}||^2 \approx 13786.1$

6: (a)

t	.6203	1.062	1.625	2.158
$\ln y$	3.524	3.081	**2.911**	**2.684**
t	3.147	8.259	8.931	9.519
$\ln y$	**2.293**	**1.036**	**1.106**	**.7467**

(b) $f(t) = 3.314 - 0.2662t$ (c) $a \approx 27.49$
(d) $y(t) = 27.49e^{-0.2662t}$

Section 7.2

1d: no

1f: yes

2b: SageMathCell 143

(i) $M^2 = \begin{bmatrix} 0.803971 & 0.175122 \\ 0.196029 & 0.824878 \end{bmatrix}$ and $M^3 = \begin{bmatrix} 0.208450997 & 0.707128254 \\ 0.791549003 & 0.292871746 \end{bmatrix}$ (ii) 0.947, 0.196029, 0.791549003 (iii) $\frac{1}{2}0.791549003 + \frac{1}{2}0.292871746 = 0.5422103745$ (iv) $\mathbf{v}_1 = \begin{bmatrix} 0.328171 \\ 0.671829 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0.585760397 \\ 0.414239603 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 0.381492005179 \\ 0.618507994821 \end{bmatrix}$ (v) 0.618507994821

2e: SageMathCell 144

(i) $M^2 = \begin{bmatrix} 0.2015 & 0.281044 & 0.268564 \\ 0.52332 & 0.484799 & 0.54607 \\ 0.275172 & 0.234157 & 0.185366 \end{bmatrix}$, $M^3 = \begin{bmatrix} 0.261059 & 0.252708 & 0.264622 \\ 0.529767 & 0.498964 & 0.508049 \\ 0.209172 & 0.248326 & 0.227327 \end{bmatrix}$ (ii) 0.21, 0.52332, 0.52976784 (iii) $\frac{1}{3}0.5297678 + \frac{1}{3}0.49896412 + \frac{1}{3}0.5080499 = 0.512260651$ (iv) $\mathbf{v}_1 = \begin{bmatrix} 0.355 \\ 0.637 \\ 0.008 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0.281044 \\ 0.484799 \\ 0.234157 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 0.25270888 \\ 0.498964123 \\ 0.248326997 \end{bmatrix}$ (v) 0.498964123

3: Because the have the same characteristic equations. Can you prove it? See crumpet 32.

4b: $\text{span}\left\{\begin{bmatrix} 846 \\ 947 \end{bmatrix}\right\}$

4e: $\text{span}\left\{\begin{bmatrix} 257969 \\ 509670 \\ 233637 \end{bmatrix}\right\}$

5b: 145

$M^{32} = \begin{bmatrix} 0.472150 & 0.471552 \\ 0.527849 & 0.528447 \end{bmatrix}$; the columns of M^{32} are approximately $\frac{1}{1793}$ times the answer from question 4b: $\frac{1}{1793}\begin{bmatrix} 846 \\ 947 \end{bmatrix}$. In other words, the ratio of the entries in each column of M^{32}, $0.47 : 0.53$, is approximately equal to the ratio of the entries in the eigenvector, $846 : 947$ (so the vectors are approximately multiples of one another). The columns of M^{32} are (nearly) in the eigenspace of M.

5e: 146

$M^{32} = \begin{bmatrix} 0.257640 & 0.257640 & 0.257640 \\ 0.509020 & 0.509020 & 0.509020 \\ 0.233339 & 0.233339 & 0.233339 \end{bmatrix}$; the columns of M^{32} are approximately $\frac{1}{1000}$ times the answer from question 4b: $\frac{1}{1000}\begin{bmatrix} 257969 \\ 509670 \\ 233637 \end{bmatrix}$. In other words, the ratio of the entries in each column of M^{32}, $0.257 : 0.509 : 0.233$, is approximately equal to the ratio of the entries in the eigenvector, $257969 : 509670 : 233637$ (so the vectors are approximately multiples of one another). The columns of M^{32} are (nearly) in the eigenspace of M.

6b: $\frac{1}{1793}\begin{bmatrix} 846 \\ 947 \end{bmatrix} \approx \begin{bmatrix} 0.471834913552705 \\ 0.528165086447295 \end{bmatrix}$

6e: $\frac{1}{1001276}\begin{bmatrix} 257969 \\ 509670 \\ 233637 \end{bmatrix} \approx \begin{bmatrix} 0.257640251039673 \\ 0.509020489854945 \\ 0.233339259105381 \end{bmatrix}$

7: (a) $\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & 1 \end{bmatrix}$ (b) 147 $0, \frac{1}{9}, \frac{8}{27}, \frac{55330}{59049} \approx 0.937$ (c) if the game can be won at all, it will eventually end

9: 148 (a) $M = \begin{bmatrix} .375 & .214 & .326 \\ .35 & .286 & .326 \\ .275 & .5 & .348 \end{bmatrix}$ (b) the consumption of each sector in dollars (c) $M\mathbf{v} = \begin{bmatrix} 26.706 \\ 30.560 \\ 42.734 \end{bmatrix}$; the farming sector consumes (26.706) more than it produces (10); the building sector produces (57) more than it consumes (30.56) (c) because 1 is an eigenvalue of every transition/consumption matrix (d) Any multiple of $\begin{bmatrix} 0.814670795745254 \\ 0.855931062340110 \\ 1 \end{bmatrix}$; for the economy in part (c), where the total economy is 100 (\$100,000), the "everybody is happy" vector is $\begin{bmatrix} 30.5051385057193 \\ 32.0501185809012 \\ 37.4447429133795 \end{bmatrix}$, for example

Section 7.3

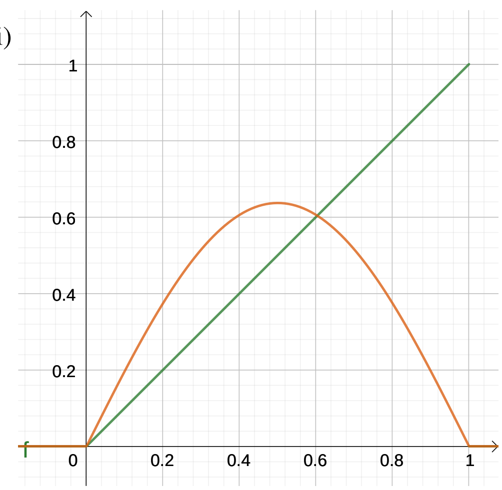
1: Verify the 10 properties of a vector space as in section 4.1

3: because $\sin\left(m\frac{\pi}{L}t\right) = \sin(0) = 0$ when $m = 0$

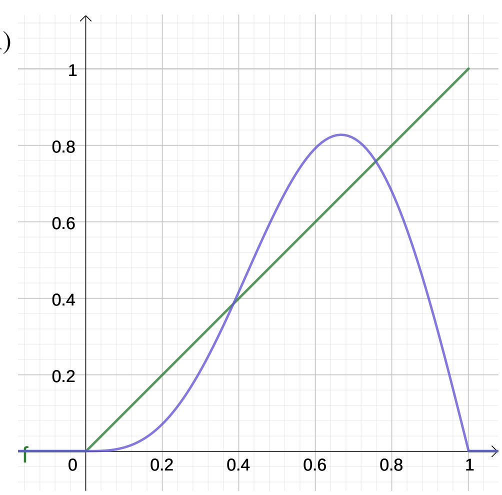
5d: $b_m = \dfrac{2m\pi}{(\ln 2)^2 + m^2\pi^2}\left(1 + 2(-1)^{m+1}\right)$

6d: $a_0 = \dfrac{1}{\ln 2}$, $a_m = \frac{2\ln 2}{(\ln 2)^2 + m^2\pi^2}\left(2(-1)^m - 1\right)$ for $m = 1, 2, \ldots$

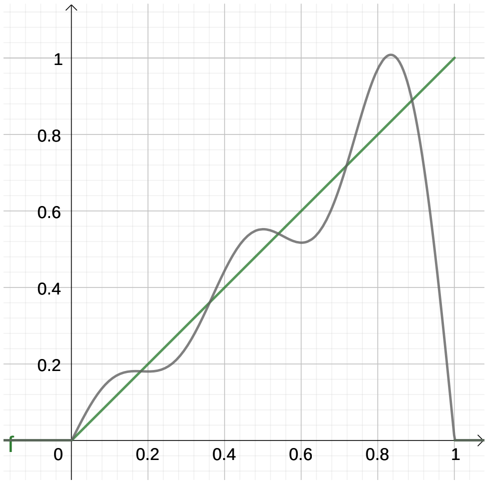
8b: (i)



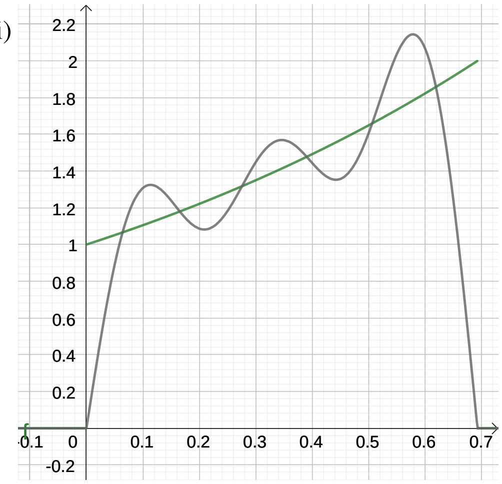
(ii)



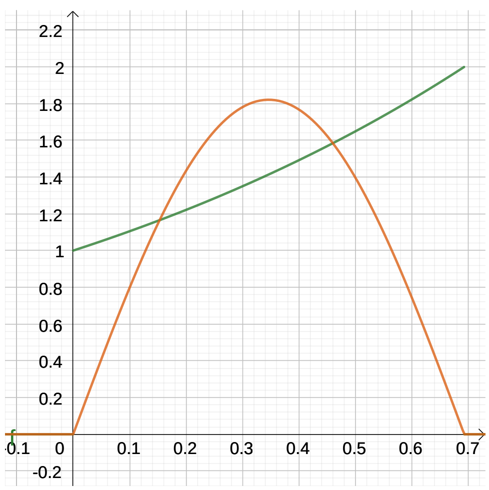
(iii)



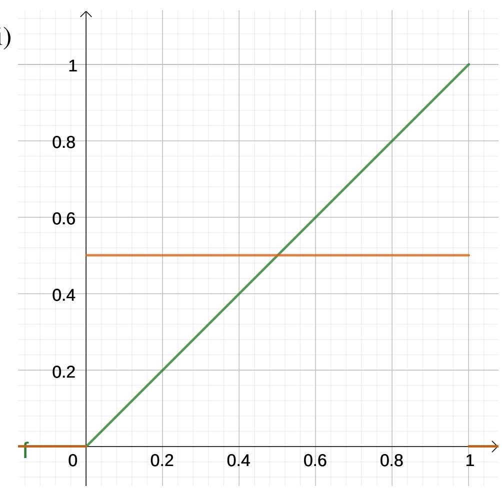
(iii)



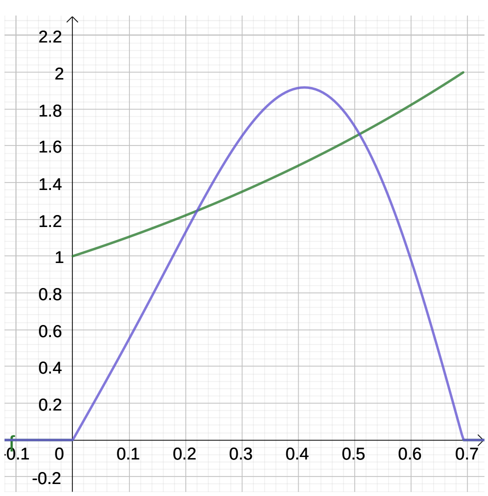
8d: (i)



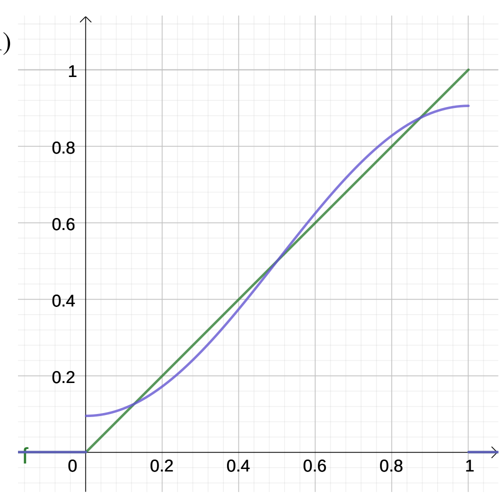
9b: (i)



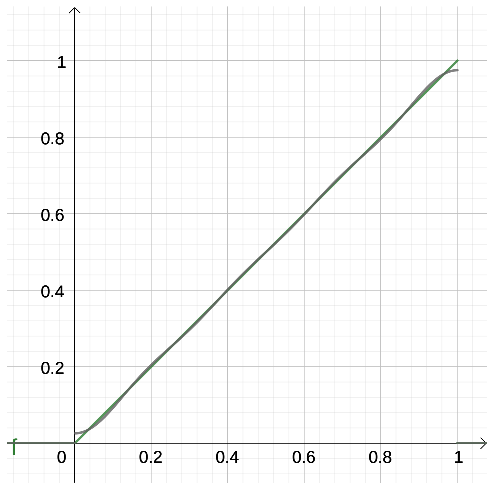
(ii)



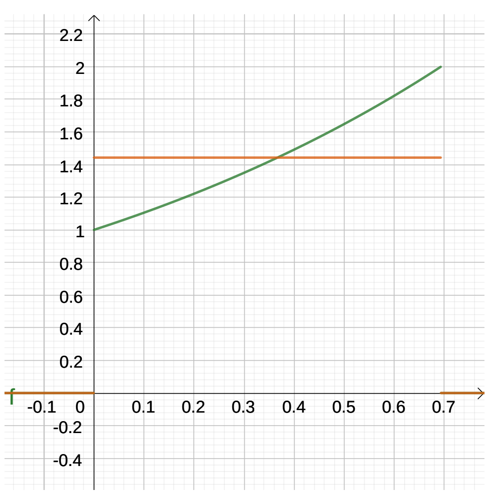
(ii)



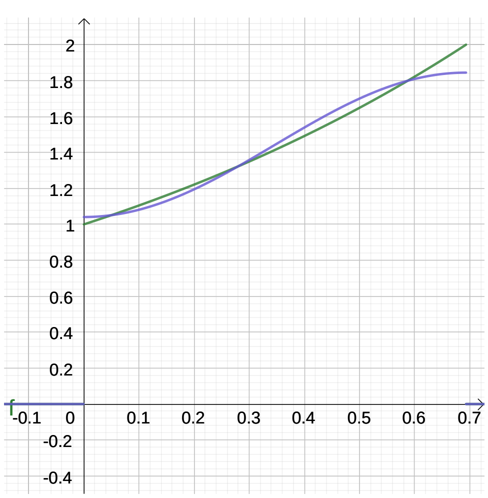
(iii)



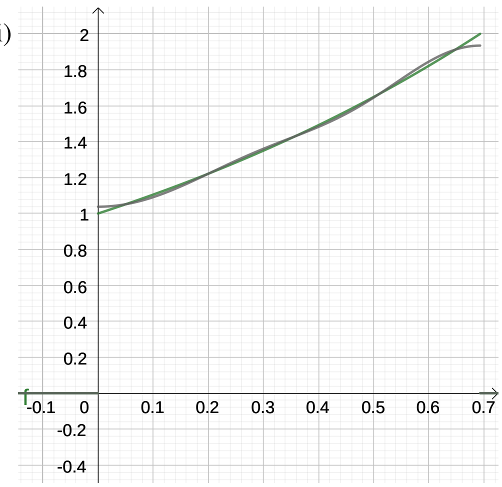
9d: (i)



(ii)



(iii)



10a: All four graphs are the same, equal to $f(x) = 1$

Section 7.4

1e: $\mathbf{v}_1 = \begin{bmatrix} 3/5 \\ -7/10 \end{bmatrix} = \begin{bmatrix} 0.6 \\ -0.7 \end{bmatrix}$

$\mathbf{v}_2 = \begin{bmatrix} 153/50 \\ 63/20 \end{bmatrix} = \begin{bmatrix} 3.06 \\ 3.15 \end{bmatrix}$

$\mathbf{v}_3 = \begin{bmatrix} 21/20 \\ 119/1000 \end{bmatrix} = \begin{bmatrix} 1.05 \\ 0.119 \end{bmatrix}$

$\mathbf{v}_4 = \begin{bmatrix} 12429/5000 \\ 22743/10000 \end{bmatrix} = \begin{bmatrix} 2.4858 \\ 2.2743 \end{bmatrix}$

1j: $\mathbf{v}_1 = \begin{bmatrix} 3/5 \\ 3/8 \\ -7/10 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.375 \\ -0.7 \end{bmatrix}$

$\mathbf{v}_2 = \begin{bmatrix} 359/80 \\ -1791/400 \\ 1171/400 \end{bmatrix} = \begin{bmatrix} 4.4875 \\ -4.4775 \\ 2.9275 \end{bmatrix}$

$\mathbf{v}_3 = \begin{bmatrix} 27199/4000 \\ -29553/4000 \\ 20751/4000 \end{bmatrix} = \begin{bmatrix} 6.79975 \\ -7.38825 \\ 5.18775 \end{bmatrix}$

$\mathbf{v}_4 = \begin{bmatrix} 319787/40000 \\ -355527/40000 \\ 254891/40000 \end{bmatrix} = \begin{bmatrix} 7.994675 \\ -8.888175 \\ 6.372275 \end{bmatrix}$

1n: $\mathbf{v}_1 = \begin{bmatrix} 1/7 \\ 2/7 \\ 3/7 \end{bmatrix} \approx \begin{bmatrix} 0.1428 \\ 0.2857 \\ 0.4285 \end{bmatrix}$

$\mathbf{v}_2 = \begin{bmatrix} 1481/196 \\ -431/196 \\ 603/196 \end{bmatrix} \approx \begin{bmatrix} 7.556 \\ -2.198 \\ 3.076 \end{bmatrix}$

$\mathbf{v}_3 = \begin{bmatrix} 1473/112 \\ -3021/784 \\ 3513/784 \end{bmatrix} \approx \begin{bmatrix} 13.15 \\ -3.853 \\ 4.480 \end{bmatrix}$

$$\mathbf{v}_4 = \begin{bmatrix} 60217/3136 \\ -17599/3136 \\ 17811/3136 \end{bmatrix} \approx \begin{bmatrix} 19.20 \\ -5.611 \\ 5.679 \end{bmatrix}$$

2e: $\mathbf{x} = \begin{bmatrix} 354/187 \\ 259/187 \end{bmatrix} \approx \begin{bmatrix} 1.893 \\ 1.385 \end{bmatrix}$

2j: $\mathbf{x} = \begin{bmatrix} 13247/1440 \\ -499/48 \\ 10907/1440 \end{bmatrix} \approx \begin{bmatrix} 9.199 \\ -10.39 \\ 7.574 \end{bmatrix}$

2n: none

3a: $-\frac{7}{6}, 1$

3e: $-\frac{7}{10}, -\frac{1}{10}$

3j: $\frac{1}{10}, \frac{1}{5}, \frac{1}{2}$

3n: $-\frac{1}{4}, \frac{3}{4}, 1$

4e: yes

4j: yes

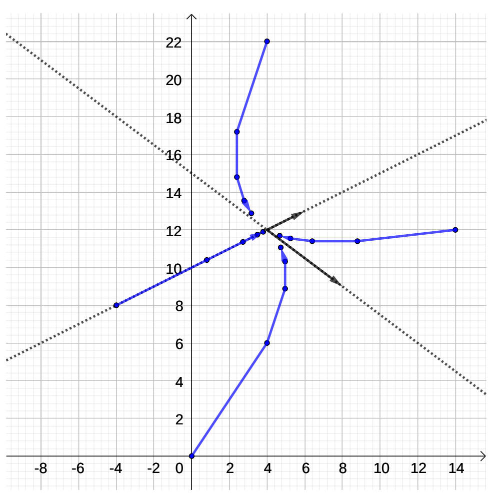
4n: no

5e: approaches attractor

5n: tends toward infinity

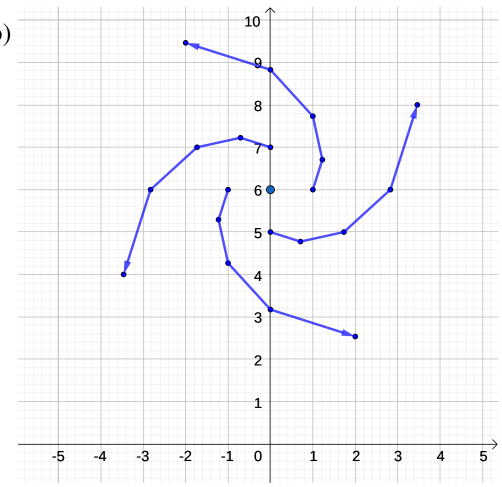
6: (a) the eigenvalues of M are $\frac{2}{5}$ and $\frac{3}{5}$ so the spectral radius is $\frac{3}{5}$, which is less than 1

(b)



9: (a) the eigenvalues of M are $\frac{\sqrt{2}}{2}(\sqrt{3}+i)$ and $\frac{\sqrt{2}}{2}(\sqrt{3}+i)$, which have the same magnitude: $\sqrt{\frac{\sqrt{2}}{2}(\sqrt{3}+i)\cdot\frac{\sqrt{2}}{2}(\sqrt{3}-i)} = \sqrt{\frac{1}{2}(3+1)} = \sqrt{2}$, which is greater than 1

(b)



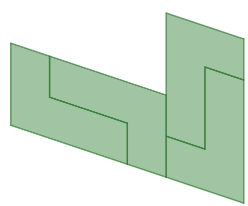
Section 7.5

1d: $\mathbf{f}(\mathbf{x}) = \frac{1}{2}\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

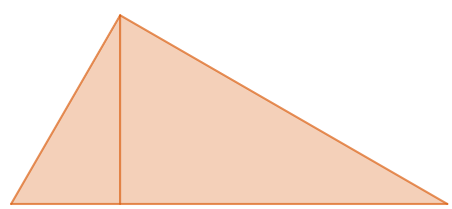
1f: $\mathbf{f}(\mathbf{x}) = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 2\sqrt{3} \end{bmatrix}$

1h: $\mathbf{f}(\mathbf{x}) = \frac{1}{2}\begin{bmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

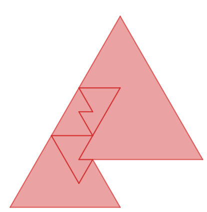
2b:



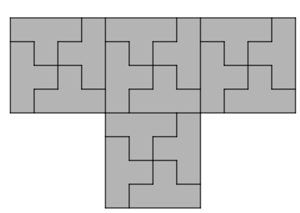
2d:



2f:



2h:

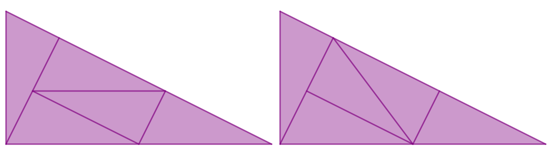


3b: Two copies of this shape can be fitted together to form a parallelogram (see the solution to exercise 2b above). Such parallelograms can be pieced together to tessellate the plane, thereby tessellating the plane with this shape.

3d: All triangles tessellate the plane. Two congruent triangles can always be put together to form a parallelogram. Such parallelograms can be pieced together to tessellate the plane, thereby tessellating the plane with the triangle.

3h: Four copies of this shape can be fitted together to form a parallelogram (see the solution to exercise 2h above). Such parallelograms can be pieced together to tessellate the plane, thereby tessellating the plane with this shape.

6:



7b: There are 5 congruent parts, so we must have $5s^2 = 1$ and therefore the scale factor is $\frac{1}{\sqrt{5}}$ for each part.

9b:
$$\left\{T_1(\mathbf{x}) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 1 \end{bmatrix},\right.$$
$$T_2(\mathbf{x}) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{x} + \begin{bmatrix} 2 \\ 0 \end{bmatrix},$$
$$T_3(\mathbf{x}) = \begin{bmatrix} -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{x} + \begin{bmatrix} 2 \\ 0 \end{bmatrix},$$
$$\left.T_4(\mathbf{x}) = \begin{bmatrix} 0 & \frac{1}{2} \\ -\frac{1}{2} & 0 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 2 \end{bmatrix}\right\}$$

9f:
$$\left\{T_1(\mathbf{x}) = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}\mathbf{x},\right.$$
$$\left.T_2(\mathbf{x}) = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}\mathbf{x} + \begin{bmatrix} 5 \\ 0 \end{bmatrix}\right\}$$

9h: Translations will vary based on placement of the axes. Placing the origin at the lower left corner of the rep-tile with scaling so that the bottom side is 2 units long, the IFS is
$$\left\{T_1(\mathbf{x}) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{x},\right.$$
$$T_2(\mathbf{x}) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$
$$\left.T_3(\mathbf{x}) = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}\mathbf{x} + \begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$$

9k: Translations will vary based on placement of the axes. Placing the origin at the center of the rep-tile with scaling so that the distance between nearest centers of the parts is 1 unit, the IFS is
$$\left\{T_1(\mathbf{x}) = \begin{bmatrix} -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{x},\right.$$
$$T_2(\mathbf{x}) = \begin{bmatrix} -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{x} + \begin{bmatrix} -1 \\ 0 \end{bmatrix},$$
$$T_3(\mathbf{x}) = \begin{bmatrix} -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{x} + \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix},$$
$$\left.T_4(\mathbf{x}) = \begin{bmatrix} -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{x} + \begin{bmatrix} \frac{1}{2} \\ -\frac{\sqrt{3}}{2} \end{bmatrix}\right\}$$

10: In the form of the required compositions, the four transormations (see the solution of exercise 9b) are

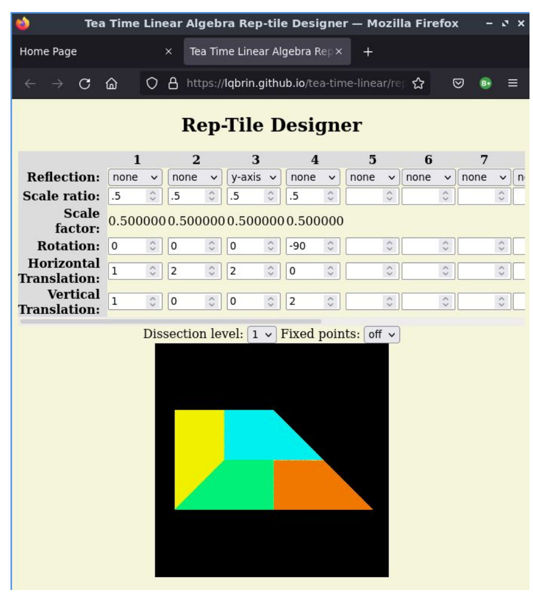
none, $\frac{1}{2}$, 0, 1, 1

none, $\frac{1}{2}$, 0, 2, 0

y-axis, $\frac{1}{2}$, 0, 2, 0

none, $\frac{1}{2}$, −90, 0, 2

Screenshot of the Rep-Tile Designer with these parameters:

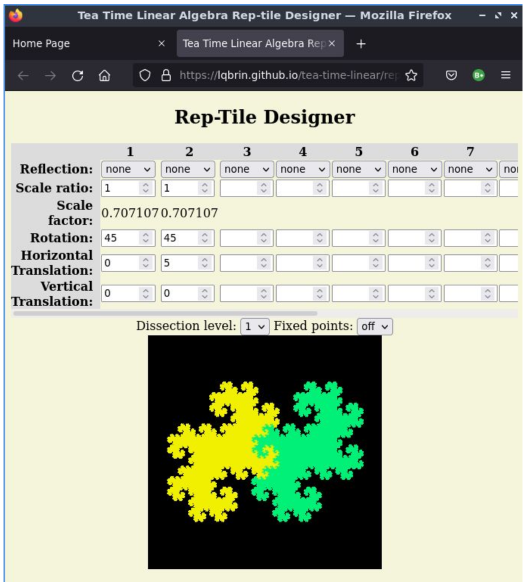


10: In the form of the required compositions, the four transormations (see the solution of exercise 9f) are

none, $\frac{1}{\sqrt{2}}$, 45, 0, 0

none, $\frac{1}{\sqrt{2}}$, 45, 5, 0

Screenshot of the Rep-Tile Designer with these parameters:



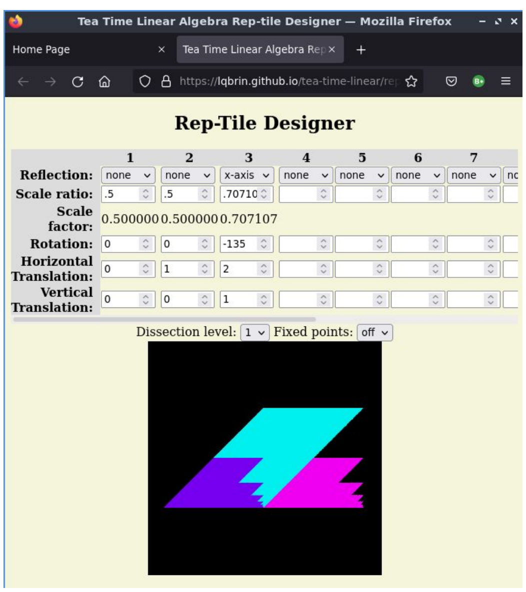
10: In the form of the required compositions, the four transormations (see the solution of exercise 9h) are

none, $\frac{1}{2}$, 0, 0, 0

none, $\frac{1}{2}$, 0, 1, 0

x-axis, $\frac{1}{\sqrt{2}}$, -135, 2, 1

Screenshot of the Rep-Tile Designer with these parameters:



10: In the form of the required compositions, the four transormations (see the solution of exercise 9k) are

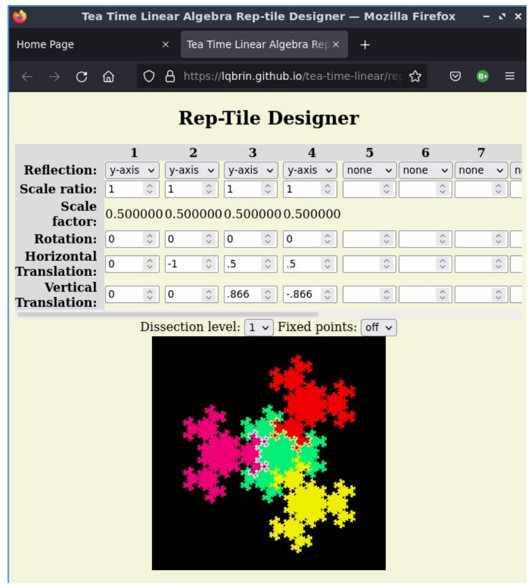
y-axis, $\frac{1}{2}$, 0, 0, 0

y-axis, $\frac{1}{2}$, 0, -1, 0

y-axis, $\frac{1}{2}$, 0, $\frac{1}{2}$, $\frac{\sqrt{3}}{2}$

y-axis, $\frac{1}{2}$, 0, $\frac{1}{2}$, $-\frac{\sqrt{3}}{2}$

Screenshot of the Rep-Tile Designer with these parameters:



Index

addition
- matrix, *see* matrix, addition
- property of equality, 79, **82**
- table, *72*
- vector, *see* vector, addition

adjugate, 35, **37**
affine transformation, *see* transformation, affine
algebraic properties of matrices, *see* matrix, algebraic properties
analysis, *71*
area, definition of, *210*
ASCII, 36, 37
attractor, 261, **265**
augmented matrix, *see* matrix, augmented

basic variable, *see* variable, basic
basis, 125, **127**
- change of, 167–169, **170**
- orthogonal, **179**, 219
- standard, **127**, 127

best approximation, ***219***, **221**, 228, 248
binomial coefficient, *96*

Cardano, Gerolamo, *201*
Cartesian product, 130, **132**
Cauchy, Augustin-Louis, *71*
change of basis, *see* basis, change of
characteristic equation, **44**, 44
characteristic polynomial, **44**, 44
characterization of matrices, *see* matrix, characterization
codomain, 130, **132**
coefficient, 28, **30**
- binomial, *see* binomial coefficient
- Fourier, *see* Fourier, coefficient
- matrix, *see* matrix, coefficient

cofactor, 27, **30**
column space, 159, **162**
consistent linear system, *see* linear system, consistent
coordinate vector, 167–169, **169**
cryptography, 35
cubic formula, *201*

decomposition, **199**
Dedekind, Richard, *71*
del Ferro, Scipione, *201*
dependence, linear, *see* linear dependence
determinant, 27–30, **30**
- algorithm, 106, **107**
- and area, 210, **214**
- and eigenvalues, ***211***, 211, **214**
- and volume, 213, **214**
- by expansion, *97*, 97, **103**
- of inverse, 110, **112**
- of lower triangular matrix, 92, **93**
- of product, ***110***, **112**
- of replacement matrix, 96, 102, **103**
- of scale matrix, 96, 102, **103**
- of swap matrix, 96, 97, *98*, 102, **103**
- of transpose, 102, **103**
- of upper triangular matrix, **93, 94**

diagonalization, 187
dilation, 270, **273**
dimension, 126, **127**
discrete dynamical system, 256–264, **264**
discriminant, 27
distance, **23**, **154**, 154
domain, 130, **132**
dominant eigenvalue, *see* eigenvalue, dominant
dot product, 14, **16**, 21, 153

echelon form
- reduced row, 55, **60**, **67**, *160*
- row, 55, **60**, 62–63

eigenpair, 42–44, **44**, 75
eigenspace, 159, **162**
eigenvalue, 42, **44**, 75
- dominant, 204, **205**, *264*

eigenvector, 42, **44**, 75
- by row reduction, 111, **112**

elementary matrix, 51, **52**, **103**, 140, **144**, 210
- determinant, 96–103
- inverse, 95–96

elementary row operation, 51, **52**, **163**
- replace, 51, **52**, 55

scale, 51, **52**, 55
swap, 51, **52**, 55
equivalent statements, 87, **88**
existence, *see* linear system, existence and uniqueness of solutions

factorization, **199**
LU, 195–198, **199**
diagonal, 187
triangular, **214**
Fejér, *249*
Ferrari, Lodovico, *201*
field, *10*
fixed point, 261, **265**
Fourier
analysis, **254**
coefficient, 252, **254**
cosine series, 252, **254**
series, 246–253, **253**
sine series, 252, **253**
free variable, *see* variable, free
function, 130, **132**

Gardner, Martin, *270*
Gershgorin circle theorem, *239*
Golomb, Solomon, 209, *270*

harmonic, 252, **254**
Hill, Lester S., *37*
homogeneous linear system, *see* linear system, homogeneous
Hutchinson's theorem, *271*, 271, **273**
hyperspace, 213

identity, *see* matrix, identity
IFS, *see* iterated function system
image, **132**, 132
set, 210, **213**
inconsistent linear system, *see* linear system, inconsistent
independence
linear, *see* linear independence
induction, *71*, 92, **93**
inheritance, 29
initial condition, 259, **264**
inner product, 153, **154**, 248
inner product space, 153–154, **154**
instantiation, *5*
inverse function, 130, **132**, 132
inverse matrix, *see* matrix, inverse
inverse relation, 130, **132**
invertible function, **132**
invertible matrix, *see* matrix, invertible
isomorphic, **149**
isomorphism, 148–149, **149**

iterate, 261, **264**
iterated function system, **273**
iteration, 202, 261, **264**

kernel, **132**, 132

Landau, Edmund, *71*
least squares regression, *see* regression, linear
linear combination, *28*, 28, **30**
nontrivial, 86, **88**
linear dependence, 86, **88**, ***126***, **127**, 160
linear equation, 49, **52**
solution, 49, **52**
linear independence, 85–87, **88**, 125, **179**
linear system, 49, **52**
characterization of solutions, **162**, ***162***
consistent, 63, **67**
existence and uniqueness of solutions, ***65***, **67**
general solution, *64*, 162
homogeneous, 59, **60**
homogeneous solution, *64*
inconsistent, 63, **67**
matrix form, 80–82, **82**
nonhomogeneous, 59, **60**
particular solution, *64*, 162
solution, 49, **52**
linear transformation, 131–132, **133**, **143**, 210
characterization, 143, **144**
geometric interpretation, 137–143
long term behavior, *264*, 261–264, **265**
lower triangular matrix, *see* matrix, lower triangular
LU factorization, *see* factorization, *LU*

main diagonal, 33, **38**
map, 131, **132**
mapping, 131, **132**
contraction, 273
Markov chain, *239*, 237–242, **242**
matrix, *3*, 3–4, **5**
addition, 9, **11**
algebraic properties, ***73***, ***74***, 71–75, **75**
augmented, 59, **60**
characterization of, ***86***, **88**, ***91***, 86–93, **93**, ***106***, 106–107
coding, 35
coefficient, 59, **60**
column, 13, **16**
diagonalizable, **188**
division, *see* matrix, inverse
elementary, *see* elementary matrix)
entry, 3, *4*, **5**
equality, 4, **5**
identity, 33, **37**
inverse, *34*, 33–37, **38**
inverse algorithm, 110, *111*, **112**

invertible, 35, **38**
lower triangular, 92, **93**
multiplication, 13–15, **16**, 80, 81, **82**
nonnegative, *239*
notation, **5**
permutation, 197, **199**
positive, *239*
powers of, 186, **188**
product, *see* matrix, multiplication
projection, 142, **144**
row, 13, **15**
similar, **188**
size, 3, **5**
square, 27, **30**
standard, ***139***, 138–139, **144**
stochastic, *239*
subtraction, 9, **11**
transition, 241, **242**
transpose, 14, **16**
upper triangular, 92, **93**
zero, 73, **75**
multiplication
matrix, *see* matrix, multiplication
property of equality, 79, **82**
scalar, 10, **11**

nonhomogeneous linear system, *see* linear system, nonhomogeneous
nontrivial solution, *see* solution, nontrivial
norm, **154**, 154
normal equations, *232*, 232–233, **233**
normalize, **179**
notation
function, 130
matrix, *see* matrix, notation
vector space, *see* vector space, notation
null space, 159, **162**
nullity, 159, ***162***, **163**

one-to-one, 148, **149**
onto, 148, **149**
operator
binary, *9*, **10**
orbit, 261, **264**
orthogonal, 22, **23**, **154**, 154, 218, **221**
basis, *see* basis, orthogonal
complement, 218, **221**
decomposition, ***219***, **221**
projection, 176, **179**, 218, **221**
set, 178, **179**
orthogonalization, 176–178, **179**, 232
orthonormal set, 178, **179**

parametric vector form, **67**
partial pivoting, 197, **199**
Peano, Giuseppe, *71*
permutation matrix, *see* matrix, permutation
perpendicular vectors, 21
pivot, 55, **60**
column, 55, **60**
position, 55, **60**, 63
power method, *204*, 201–205, **205**
preimage, **132**, 132
product
Cartesian, *see* Cartesian product
dot, *see* dot product
inner, *see* inner product
matrix, *see* matrix, multiplication
scalar, *see* multiplication, scalar
projection, 142
orthogonal, *see* orthogonal, projection
projection matrix, *see* matrix, projection
proof technique
contraposition, *87*
equivalent statements, 87
induction, 92, **93**
real numbers are equal, *106*

quartic formula, *201*
quintic polynomial, 201

range, 132, **133**
rank, 159, ***162***, **163**
recurrence, 259, **264**
reduced row echelon form, *see* echelon form, reduced row
regression
linear, 227–233, **233**
multilinear, 231, **233**
multiple linear, 231, **233**
relation, 130, **132**
recurrence, **264**
rep-tile, 268–272, **273**
repeller, **265**
ring, *15*
row echelon form, *see* echelon form, row
row reduction, 55–59, **60**
automated, *55*

SageMath
`charpoly()`, 44
coercion, 23
`determinant`, 30
dot product, 17
`echelon_form()`, 66
`eigenvalues()`, 44
`eigenvectors_right()`, 44
`inverse()`, 38
magnitude, 24
`matrix()`, 6

nested statement, *7*
`norm()`, 24
operators, 11
`rref()`, 66
SageCell screenshot, 6
submatrix, 6
`transpose()`, 16
`vector()`, 23
scalar, 10, **11**
product, *see* multiplication, scalar
set
bounded, **273**
closed, **273**
compact, **273**
image, *see* image, set
orthogonal, *see* orthogonal, set
orthonormal, *see* orthonormal set
solution, *see* solution, set
spanning, *see* spanning set
similarity, ***186***, 186–188, **188**
similitude, 270, **273**
solution
least squares, **233**
linear equation, *see* linear equation, solution
linear system, *see* linear system, solution
nontrivial, **82**, 82, 86, **88**
set, 62–66
trivial, 86, **88**
solution space, 159–162
span, 119, **120**, 125, **127**
spanning set, 125, **127**
spectral radius, 263, **265**
standard basis, *see* basis, standard
standard matrix, *see* matrix, standard
submatrix, 4, **5**
subspace, 119, **120**
subtraction
matrix, *see* matrix, subtraction
vector, *see* vector, subtraction
sum of squared errors, 228, **233**

Tartaglia, Niccolò, *201*
transformation, 131, **132**
affine, 213, **214**, **264**
rigid, 270, **273**
transpose, *see* matrix, transpose
triangularization, *212*, **214**
trivial vector space, *see* vector space, trivial

uniqueness, *see* linear system, existence and uniqueness of solutions
unit vector, *see* vector, unit
upper triangular matrix, *see* matrix, upper triangular

variable
basic, 65, **67**
free, 65, **67**
vector, 14, **16**, 118, **120**
addition, 21, **22**
column, 15, **16**
coordinate, *see* coordinate vector
geometric interpretation, 20, **22**, **143**
magnitude, 21, **22**
orthogonality, *see* orthogonal
representation, 147
row, *15*, **16**
subtraction, 22, **23**
unit, 178, **179**
zero, 43, **44**
vector space, **119**, 117–120
notation, 120
trivial, 126, **127**

Yanghui triangle, *49*

zero matrix, *see* matrix, zero

Bibliography

[1] C. Adam Berrey. Making absolute population estimates in the intermediate area using the area and density of ceramic sherd scatters: An application of regression analysis. *Journal of Archaeological Science*, 97:147–158.

[2] Christoph Bandt. Self-similar sets 5. integer matrices and fractal tilings of $\mathbb{R}^n$. *Proceedings of the American Mathematical Society*, 112(2):549–562, June 1991.

[3] Jamie B. Barker, Matthew J. Slater, Geoff Pugh, Stephen D. Mellalieu, Paul J. McCarthy, Marc V. Jones, and Aidan Moran. The effectiveness of psychological skills training and behavioral interventions in sport using single-case designs: A meta regression analysis of the peer-reviewed studies. *Psychology of Sport and Exercise*, 51:101746, 2020.

[4] Richard L. Burden, J. Douglas Faires, and Annette M. Burden. *Numerical Analysis*. Cengage Learning, 10th edition, 2016.

[5] D. Caroline Coile and Margaret H. Bonham. *Why Do Dogs Like Balls?: More Than 200 Canine Quirks, Curiosities, and Conundrums Revealed*. Sterling Publishing Company, Inc, 2008.

[6] Ángela García, Ofélia Anjos, Carla Iglesias, Helena Pereira, Javier Martínez, and Javier Taboada. Prediction of mechanical strength of cork under compression using machine learning techniques. *Materials & Design*, 82:304–311, 2015.

[7] Martin Gardner. Mathematical games. *Scientific American*, 199(5):136–144, 1958.

[8] Tita G.E. and Radil S.M. Spatial regression models in criminology: Modeling social processes in the spatial weights matrix. 2010.

[9] Kurt Gödel. Über formal unentscheidbare sätze der principia mathematica und verwandter systeme i. *Monatsh. f. Mathematik und Physik*, 38:173–198, 1931. https://doi.org/10.1007/BF01700692.

[10] Solomon W. Golomb. Replicating figures in the plane. *The Mathematical Gazette*, 48(366):403–412, Dec 1964.

[11] Roger Hart. *The Chinese Roots of Linear Algebra*. Johns Hopkins University Press, 2011.

[12] Lester S. Hill. Cryptography in an algebraic alphabet. *Mathematical Monthly*, 36(6):306–312, June-July 1929.

[13] William L. Hosch. Pascal's triangle, August 2009. Encyclopedia Britannica. https://www.britannica.com/science/Pascals-triangle. Accessed January 29, 2021.

[14] John E. Hutchinson. Fractals and self similarity. *Indiana University Mathematics Journal*, 30(5):713–747, 1981.

[15] Edmund Landau. *Foundations of Analysis: The Arithmetic of Whole, Rational, Irrational and Complex Numbers*. Chelsea Publishing Company, third edition, 1966.

[16] Mariusz Malinowski and Lidia Jabłońska-Porzuczek. Female activity and education levels in selected european union countries. *Research in Economics*, 74(2):153–173, 2020.

[17] Elizabeth S. Meckes and Mark W. Meckes. *Linear Algebra*. Cambridge University Press, 2018.

[18] BBC News Online. Technology | key computer coding creator dies, June 2004. http://news.bbc.co.uk/2/hi/technology/3838845.stm. Accessed January 27, 2021.

[19] K.B. Payne, W.R. Langbauer, and E.M. Thomas. Infrasonic calls of the asian elephant (*Elephas maximus*). *Behav Ecol Sociobiol*, 18:297–301, 1986.

[20] Ioseph Peano. *Arithmetices Principia: Nova Methodo Exposita*. Augustae Taurinorum: Fratres Bocca, 1889.

[21] Mark A. Rose. Predicting daily maximum temperatures using linear regression and geopotential thickness forecasts, unknown. National Weather Service. https://www.weather.gov/ohx/predictingdailymaxtemps. Accessed April 5, 2021.

[22] Brian S. Thomson, Judith B. Bruckner, and Andrew M. Bruckner. *Elementary Real Analysis*. ClassicalRealAnalysis.com, second edition edition, 2008.

[23] B. Wuyts, M. Maenhoudt, S. Libbrecht, Z.X. Gao, E. Osquiguil, V.V. Moshchalkov, and Y. Bruynseraede. Linear relation between the hall angle slope and the carrier density in yba2cu3ox films. *Physica C: Superconductivity*, 235-240:1369–1370, 1994.