APPLIED LINEAR ALGEBRA AND MATRIX ANALYSIS

Thomas S. Shores

Author address:

Contents

Preface

This book is about matrix and linear algebra, and their applications. For many students the tools of matrix and linear algebra will be as fundamental in their professional work as the tools of calculus; thus it is important to ensure that students appreciate the utility and beauty of these subjects, as well as understand the mechanics. One way to do so is to show how concepts of matrix and linear algebra make concrete problems workable. To this end, applied mathematics and mathematical modeling ought to have an important role in an introductory treatment of linear algebra.

One of the features of this book is that we weave significant motivating examples into the fabric of the text. Needless to say, I hope that instructors will not omit this material; that would be a missed opportunity for linear algebra! The text has a strong orientation towards numerical computation and applied mathematics, which means that matrix analysis plays a central role. All three of the basic components of linear algebra – theory, computation and applications – receive their due. The proper balance of these components will give a diverse audience of physical science, social science, statistics, engineering and math students the tools they need as well as the motivation to acquire these tools. Another feature of this text is an emphasis on linear algebra as an experimental science; this emphasis is to be found in certain examples, computer exercises and projects. Contemporary mathematical software makes an ideal "lab" for mathematical experimentation. At the same time, this text is independent of specific hardware and software platforms. Applications and ideas should play center stage, not software.

This book is designed for an introductory course in matrix and linear algebra. It is assumed that the student has had some exposure to calculus. Here are some of its main goals:

- To provide a balanced blend of applications, theory and computation which emphasizes their interdependence.
- To assist those who wish to incorporate mathematical experimentation through computer technology into the class. Each chapter has an optional section on computational notes and projects and computer exercises sprinkled throughout. The student should use the locally available tools to carry out the experiments suggested in the project and use the word processing capabilities of the computer system to create small reports on his/her results. In this way they gain experience in the use of the computer as a mathematical tool. One can also envision reports on a grander scale as mathematical "term papers." I have made such assignments in some of my own classes with delightful results. A few major report topics are included in the text.

- To help students to think precisely and express their thoughts clearly. Requiring written reports is one vehicle for teaching good expression of mathematical ideas. The projects given in this text provide material for such reports.
- To encourage cooperative learning. Mathematics educators are becoming increasingly appreciative of this powerful mode of learning. Team projects and reports are excellent vehicles for cooperative learning.
- To promote individual learning by providing a complete and readable text. I hope that students will find the text worthy of being a permanent part of their reference library, particularly for the basic linear algebra needed for the applied mathematical sciences.

An outline of the book is as follows: Chapter 1 contains a thorough development of Gaussian elimination and an introduction to matrix notation. It would be nice to assume that the student is familiar with complex numbers, but experience has shown that this material is frequently long forgotten by many. Complex numbers and the basic language of sets are reviewed early on in Chapter 1. (The advanced part of the complex number discussion could be deferred until it is needed in Chapter 4.) In Chapter 2, basic properties of matrix and determinant algebra are developed. Special types of matrices, such as elementary and symmetric, are also introduced. About determinants: some instructors prefer not to spend too much time on them, so I have divided the treatment into two sections, one of which is marked as optional and not used in the rest of the text. Chapter 3 begins by introducing the student to the "standard" Euclidean vector spaces, both real and complex. These are the well springs for the more sophisticated ideas of linear algebra. At this point the student is introduced to the general ideas of abstract vector space, subspace and basis, but primarily in the context of the standard spaces. Chapter 4 introduces goemetrical aspects of standard vectors spaces such as norm, dot product and angle. Chapter 5 provides an introduction to eigenvalues and eigenvectors. Subsequently, general norm and inner product concepts are examined in Chapter 5. Two appendices are devoted to a table of commonly used symbols and solutions to selected exercises.

Each chapter contains a few more "optional" topics, which are independent of the non-optional sections. I say this realizing full well that one instructor's optional is another's mandatory. Optional sections cover tensor products, linear operators, operator norms, the Schur triangularization theorem and the singular value decomposition. In addition, each chapter has an optional section of computational notes and projects. I have employed the convention of marking sections and subsections that I consider optional with an asterisk. Finally, at the end of each chapter is a selection of review exercises.

There is more than enough material in this book for a one semester course. Tastes vary, so there is ample material in the text to accommodate different interests. One could increase emphasis on any one of the theoretical, applied or computational aspects of linear algebra by the appropriate selection of syllabus topics. The text is well suited to a course with a three hour lecture and lab component, but the computer related material is not mandatory. Every instructor has her/his own idea about how much time to spend on proofs, how much on examples, which sections to skip, etc.; so the amount of material covered will vary considerably. Instructors may mix and match any of the optional sections according to their own interests, since these sections are largely independent of each other. My own opinion is that the ending sections in each chapter on computational notes and projects are partly optional. While it would be very time consuming to cover them all, every instructor ought to use some part of this material. The unstarred sections form the core of the book; most of this material should be covered. There are 27 unstarred sections and 12 optional sections. I hope the optional sections come in enough flavors to please any pure, applied or computational palate.

Of course, no one shoe size fits all, so I will suggest two examples of how one might use this text for a three hour one semester course. Such a course will typically meet three times a week for fifteen weeks, for a total of 45 classes. The material of most of the the unstarred sections can be covered at a rate of about one and one half class periods per section. Thus, the core material could be covered in about 40 class periods. This leaves time for extra sections and and in-class exams. In a two semester course or a semester course of more than three hours, one could expect to cover most, if not all, of the text.

If the instructor prefers a course that emphasizes the standard Euclidean spaces, and moves at a more leisurely pace, then the core material of the first five chapters of the text are sufficient. This approach reduces the number of unstarred sections to be covered from 27 to 23.

In addition to the usual complement of pencil and paper exercises (with selected solutions in Appendix B), this text includes a number of computer related activities and topics. I employ a taxonomy for these activities which is as follows. At the lowest level are computer exercises. Just as with pencil and paper exercises, this work is intended to develop basic skills. The difference is that some computing equipment (ranging from a programmable scientific calculator to a workstation) is required to complete such exercises. At the next level are computer projects. These assignments involve ideas that extend the standard text material, possibly some experimentation and some written exposition in the form of brief project papers. These are analogous to lab projects in the physical sciences. Finally, at the top level are reports. These require a more detailed exposition of ideas, considerable experimentation – possibly open ended in scope, and a carefully written report document. Reports are comparable to "scientific term papers". They approximate the kind of activity that many students will be involved in through their professional life. I have included some of my favorite examples of all three activities in this textbook. Exercises that require computing tools contain a statement to that effect. Perhaps projects and reports I have included will be paradigms for instructors who wish to build their own project/report materials. In my own classes I expect projects to be prepared with text processing software to which my students have access in a mathematics computer lab.

Projects and reports are well suited for team efforts. Instructors should provide background materials to help the students through local system dependent issues. For example, students in my own course are assigned a computer account in the mathematics lab and required to attend an orientation that contains specific information about the available linear algebra software. When I assign a project, I usually make available a Maple or Mathematica notebook that amounts to a brief background lecture on the subject of the project and contains some of the key commands students will need to carry out the project. This helps students focus more on the mathematics of the project rather than computer issues.

Most of the computational computer tools that would be helpful in this course fall into three categories and are available for many operating systems:

- Graphing calculators with built-in matrix algebra capabilities such as the HP 28 and 48, or the TI 85 and 92. These use floating point arithmetic for system solving and matrix arithmetic. Some do eigenvalues.
- Computer algebra systems (CAS) such as Maple, Mathematica and Macsyma. These software products are fairly rich in linear algebra capabilities. They prefer symbolic calculations and exact arithmetic, but will do floating point calculations, though some coercion may be required.
- Matrix algebra systems (MAS) such as MATLAB or Octave. These software products are specifically designed to do matrix calculations in floating point arithmetic, though limited symbolic capabilities are available in the basic program. They have the most complete set of matrix commands of all categories.

In a few cases I have included in this text some software specific information for some projects, for the purpose of illustration. This is not to be construed as an endorsement or requirement of any particular software or computer. Projects may be carried out with different software tools and computer platforms. Each system has its own strengths. In various semesters I have obtained excellent results with all these platforms. Students are open to all sorts of technology in mathematics. This openness, together with the availability of inexpensive high technology tools, is changing how and what we teach in linear algebra.

I would like to thank my colleagues whose encouragement has helped me complete this project, particularly Jamie Radcliffe, Jim Lewis, Dale Mesner and John Bakula. Special thanks also go to Jackie Kohles for her excellent work on solutions to the exercises and to the students in my linear algebra courses for relentlessly tracking down errors. I would also like to thank my wife, Muriel, for an outstanding job of proofreading and editing the text.

I'm in the process of developing a linear algebra home page of material such as project notebooks, supplementary exercises, etc, that will be useful for instructors and students of this course. This site can be reached through my home page at

`http://www.math.unl.edu/~tshores/`

I welcome suggestions, corrections or comments on the site or book; both are ongoing projects. These may be sent to me at `tshores@math.unl.edu`.

CHAPTER 1

LINEAR SYSTEMS OF EQUATIONS

There are two central problems about which much of the theory of linear algebra revolves: the problem of finding all solutions to a linear system and that of finding an eigensystem for a square matrix. The latter problem will not be encountered until Chapter 4; it requires some background development and even the motivation for this problem is fairly sophisticated. By contrast the former problem is easy to understand and motivate. As a matter of fact, simple cases of this problem are a part of the high school algebra background of most of us. This chapter is all about these systems. We will address the problem of when a linear system has a solution and how to solve such a system for all of its solutions. Examples of linear systems appear in nearly every scientific discipline; we touch on a few in this chapter.

1.1. Some Examples

Here are a few elementary examples of linear systems:

EXAMPLE 1.1.1. For what values of the unknowns x and y are the following equations satisfied?

$$\begin{aligned} x + 2y &= 5 \\ 4x + y &= 6 \end{aligned}$$

SOLUTION. The first way that we were taught to solve this problem was the geometrical approach: every equation of the form $ax + by + c = 0$ represents the graph of a straight line, and conversely, every line in the xy-plane is so described. Thus, each equation above represents a line. We need only graph each of the lines, then look for the point where these lines intersect, to find the unique solution to the graph (see Figure 1.1.1). Of course, the two equations may represent the same line, in which case there are infinitely many solutions, or distinct parallel lines, in which case there are no solutions. These could be viewed as exceptional or “degenerate” cases. Normally, we expect the solution to be unique, which it is in this example.

We also learned how to solve such an equation algebraically: in the present case we may use either equation to solve for one variable, say x, and substitute the result into the other equation to obtain an equation which is easily solved for y. For example, the first equation above yields $x = 5 - 2y$ and substitution into the second yields $4(5 - 2y) + y = 6$, i.e., $-7y = -14$, so that $y = 2$. Now substitute 2 for y in the first equation and obtain that $x = 5 - 2(2) = 1$. $\square$

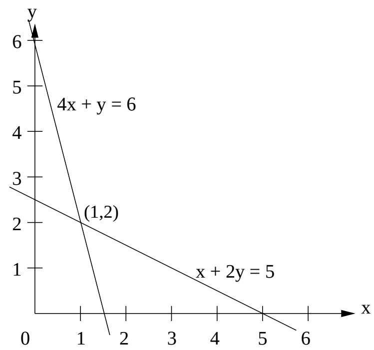


FIGURE 1.1.1. Graphical solution to Example 1.1.1.

EXAMPLE 1.1.2. For what values of the unknowns x, y and z are the following equations satisfied?

$$\begin{aligned} x + y + z &= 4 \\ 2x + 2y + 5z &= 11 \\ 4x + 6y + 8z &= 24 \end{aligned}$$

SOLUTION. The geometrical approach becomes somewhat impractical as a means of obtaining an explicit solution to our problem: graphing in three dimensions on a flat sheet of paper doesn't lead to very accurate answers! Nonetheless, the geometrical point of view is useful, for it gives us an idea of what to expect without actually solving the system of equations.

With reference to our system of three equations in three unknowns, the first fact to take note of is that each of the three equations is an instance of the general equation $ax + by + cz + d = 0$. Now we know from analytical geometry that the graph of this equation is a plane in three dimensions, and conversely every such plane is the graph of some equation of the above form. In general, two planes will intersect in a line, though there are exceptional cases of the two planes represented being identical or distinct and parallel. Hence we know the geometrical shape of the solution set to the first two equations of our system: a plane, line or point. Similarly, a line and plane will intersect in a point or, in the exceptional case that the line and plane are parallel, their intersection will be the line itself or the empty set. Hence, we know that the above system of three equations has a solution set that is either empty, a single point, a line or a plane.

Which outcome occurs with our system of equations? We need the algebraic point of view to help us calculate the solution. The matter of dealing with three equations and three unknowns is a bit trickier than the problem of two equations and unknowns. Just as with two equations and unknowns, the key idea is still to use one equation to solve for one unknown. Since we have used one equation up, what remains is two equations in the remaining unknowns. In this problem, subtract 2 times the first equation from the second and 4 times the first equation from the third to obtain the system

$$\begin{aligned} 3z &= 3 \\ 2y + 4z &= 8 \end{aligned}$$

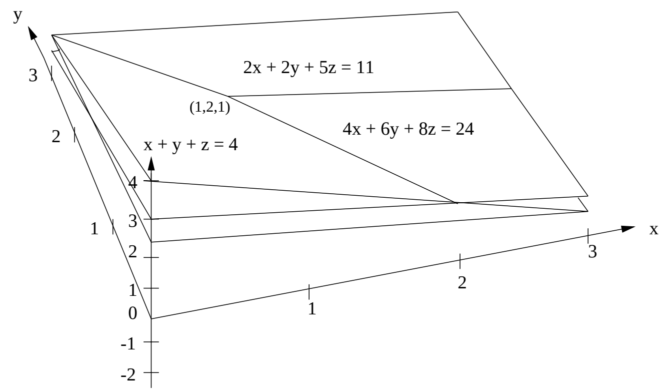


FIGURE 1.1.2. Graphical solution to Example 1.1.2.

which are easily solved to obtain $z = 1$ and $y = 2$. Now substitute into the first equation and obtain that $x = 1$. We can see that the graphical method of solution becomes impractical for systems of more than two variables, though it still tells us about the qualitative nature of the solution. This solution can be discerned roughly in Figure 1.1.2. □

Some Key Notation

Here is a formal statement of the kind of equation that we want to study in this chapter. This formulation gives us a means of dealing with the general problem later on.

DEFINITION 1.1.3. A *linear equation* in the variables $x_1, x_2, ..., x_n$ is an equation of the form

$$a_1x_1 + a_2x_2 + ... + a_nx_n = b$$

where the coefficients $a_1, a_2, ..., a_n$ and right hand side constant term b are given constants.

Of course, there are many interesting and useful nonlinear equations, such as $ax^2 + bx + c = 0$, or $x^2 + y^2 = 1$, etc. But our focus is on systems that consist solely of linear equations. In fact, our next definition gives a fancy way of describing the general linear system.

Linear Systems

DEFINITION 1.1.4. A *linear system* of m equations in the n unknowns $x_1, x_2, ..., x_n$ is a list of m equations of the form

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1j}x_j + \cdots a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2j}x_j + \cdots a_{2n}x_n & = & b_2 \\ \vdots & \vdots & \vdots \\ a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ij}x_j + \cdots a_{in}x_n & = & b_i \\ \vdots & \vdots & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mj}x_j + \cdots a_{mn}x_n & = & b_m \end{array} \tag{1.1.1}$$

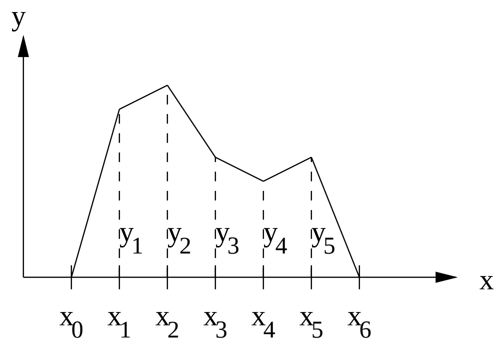


FIGURE 1.1.3. Discrete approximation to temperature function ($n = 5$).

Notice how the coefficients are indexed: in the ith row the coefficient of the jth variable, x_j, is the number a_{ij}, and the right hand side of the ith equation is b_i. This systematic way of describing the system will come in handy later,when we introduce the matrix concept.

* Examples of Modeling Problems

It is easy to get the impression that linear algebra is about the simple kinds of problems of the preceding examples. So why develop a whole subject? Next we consider two examples whose solutions will not be so apparent as the previous two examples. The real point of this chapter, as well as that of Chapters 2 and 3, is to develop algebraic and geometrical methodologies which are powerful enough to handle problems like these.

Diffusion Processes

We consider a diffusion process arising from the flow of heat through a homogeneous material substance. A basic physical observation to begin with is that heat is directly proportional to temperature. In a wide range of problems this hypothesis is true, and we shall always assume that we are modeling such a problem. Thus, we can measure the amount of heat at a point by measuring temperature since they differ by a known constant of proportionality. To fix ideas, suppose we have a rod of material of unit length, say, situated on the x-axis, for $0 \leq x \leq 1$. Suppose further that the rod is laterally insulated, but has a known internal heat source that doesn't change with time. When sufficient time passes, the temperature of the rod at each point will settle down to "steady state" values, dependent only on position x. Say the heat source is described by a function $f(x)$, $0 \leq x \leq 1$,which gives the additional temperature contribution per unit length per unit time due to the heat source at the point x. Also suppose that the left and right ends of the rod are held at fixed at temperatures y_0 and y_1.

How can we model a steady state? Imagine that the continuous rod of uniform material is divided up into a finite number of equally spaced points, called nodes, namely $x_0 = 0, x_1, ..., x_{n+1} = 1$ and that all the heat is concentrated at these points. Assume the nodes are a distance h apart. Since spacing is equal, the relation between h and n is $h = 1/(n + 1)$. Let the temperature function be $y(x)$ and let $y_i = y(x_i)$. Approximate

$y(x)$ in between nodes by connecting adjacent points (x_i, y_i) with a line segment. (See Figure 1.1.3 for a graph of the resulting approximation to $y(x)$.) We know that at the end nodes the temperature is specified: $y(x_0) = y_0$ and $y(x_{n+1}) = y_1$. By examining the process at each interior node, we can obtain the following linear equation for each interior node index $i = 1, 2, \ldots, n$ involving a constant k called the conductivity of the material. A derivation of these equations follows this example.

$$k\frac{-y_{i-1} + 2y_i - y_{i+1}}{h^2} = f(x_i)$$

or

$$-y_{i-1} + 2y_i - y_{i+1} = \frac{h^2}{k}f(x_i) \tag{1.1.2}$$

EXAMPLE 1.1.5. Suppose we have a rod of material of conductivity $k = 1$ and situated on the x-axis, for $0 \leq x \leq 1$. Suppose further that the rod is laterally insulated, but has a known internal heat source and that both the left and right ends of the rod are held at 0 degrees Fahrenheit. What are the steady state equations approximately for this problem?

SOLUTION. Follow the notation of the discussion preceding this example. Notice that in this case $x_i = ih$. Remember that y_0 and y_{n+1} are known to be 0, so the terms y_0 and y_{n+1} disappear. Thus we have from Equation 1.1.2 that there are n equations in the unknowns y_i, $i = 1, 2, \ldots, n$.

It is reasonable to expect that the smaller h is, the more accurately y_i will approximate $y(x_i)$. This is indeed the case. But consider what we are confronted with when we take $n = 5$, i.e., $h = 1/(5+1) = 1/6$, which is hardly a small value of h. The system of five equations in five unknowns becomes

$$\begin{array}{rrrrrcl}
2y_1 & -y_2 & & & & = & f(1/6)/36 \\
-y_1 & +2y_2 & -y_3 & & & = & f(2/6)/36 \\
 & -y_2 & +2y_3 & -y_4 & & = & f(3/6)/36 \\
 & & -y_3 & +2y_4 & -y_5 & = & f(4/6)/36 \\
 & & & -y_4 & +2y_5 & = & f(5/6)/36
\end{array}$$

This problem is already about as large as we would want to work by hand. The basic ideas of solving systems like this are the same as in Example 1.1.1 and 1.1.2, though for very small h, say $h = .01$, clearly we would like some help from a computer or calculator.

**Derivation of the diffusion equations.* We follow the notation that has already been developed, except that the values y_i will refer to quantity of heat rather than temperature (this will yield equations for temperature, since heat is a constant times temperature). What should happen at an interior node? The explanation requires one more experimentally observed law known as *Fourier's heat law.* It says that the flow of heat per unit length from one point to another is proportional to the rate of change in temperature with respect to distance and moves from higher temperature to lower. The constant of proportionality k is known as the *conductivity* of the material. In addition, we interpret the heat created at node x_i to be $hf(x_i)$, since f measures heat created per unit length. Count flow towards the right as positive. Thus, at node x_i the net flow per unit length from the left node and to the right node are given by

$$\begin{aligned}\text{Left flow } &= k\frac{y_i - y_{i-1}}{h}\\ \text{Right flow} &= k\frac{y_i - y_{i+1}}{h}\end{aligned}$$

Thus, in order to balance heat flowing through the ith node with heat created per unit length at this node, we should have

$$\text{Leftflow} + \text{Rightflow} = k\frac{y_i - y_{i-1}}{h} + k\frac{y_i - y_{i+1}}{h} = hf(x_i)$$

In other words,

$$k\frac{-y_{i-1} + 2y_i - y_{i+1}}{h^2} = f(x_i)$$

or

$$-y_{i-1} + 2y_i - y_{i+1} = \frac{h^2}{k}f(x_i) \tag{1.1.3}$$

Input-Output models

We are going to set up a simple model of an economy consisting of three sectors that supply each other and consumers. Suppose the three sectors are (E)nergy, (M)aterials and (S)ervices and suppose that the demands of a sector are proportional to its output. This is reasonable; if, for example, the materials sector doubled its output, one would expect its needs for energy, material and services to likewise double. Now let x, y, z be the total outputs of the sectors E,M and S respectively. We require that the economy be *closed* in the sense that everything produced in the economy is consumed by the economy. Thus, the total output of the sector E should equal the amounts consumed by all the sectors and the consumers.

EXAMPLE 1.1.6. Given the following table of demand constants of proportionality and consumer (D)emand (a fixed quantity) for the output of each service, express the closed property of the system as a system of equations.

		Consumed by			
		E	M	S	D
	E	0.2	0.3	0.1	2
Produced by	M	0.1	0.3	0.2	1
	S	0.4	0.2	0.1	3

SOLUTION. Consider how we balance the total output and demands for energy. The total output is x units. The demands from the three sectors E,M and S are, according to the table data, $0.2x$, $0.3y$ and $0.1z$, respectively. Further, consumers demand 2 units of energy. In equation form

$$x = 0.2x + 0.3y + 0.1z + 2$$

Likewise we can balance the input/output of the sectors M and S to arrive at a system of three equations in three unknowns.

$$\begin{aligned} x &= 0.2x + 0.3y + 0.1z + 2 \\ y &= 0.1x + 0.3y + 0.2z + 1 \\ z &= 0.4x + 0.2y + 0.1z + 3 \end{aligned}$$

The questions that interest economists are whether or not this system has solutions, and if so, what they are. □

Note: In some of the text exercises you will find references to “your computer system.” This may be a calculator that is required for the course or a computer system for which you are given an account. This textbook does not depend on any particular system, but certain exercises require a computational device. The abbreviation “MAS” stands for a matrix algebra system like MATLAB or Octave. Also, the shorthand “CAS” stands for a computer algebra system like Maple, Mathematica or MathCad. A few of the projects are too large for most calculators and will require a CAS or MAS.

1.1 Exercises

1. Solve the following systems algebraically.

(a) $\begin{array}{rcr} x+2y &=& 1 \\ 3x-y &=& -4 \end{array}$ (b) $\begin{array}{rcl} x-y+2z &=& 6 \\ 2x-z &=& 3 \\ y+2z &=& 0 \end{array}$ (c) $\begin{array}{rcl} x-y &=& 1 \\ 2x-y &=& 3 \\ x+y &=& 3 \end{array}$

2. Determine if the following systems of equations are linear or not. If so, put them in standard format.

(a) $\begin{array}{rcl} x+2 &=& y+z \\ 3x-y &=& 4 \end{array}$ (b) $\begin{array}{rcl} xy+2 &=& 1 \\ 2x-6 &=& y \end{array}$ (c) $\begin{array}{rcl} x+2 &=& 1 \\ x+3 &=& y^2 \end{array}$

3. Express the following systems of equations in the notation of the definition of linear systems by specifying the numbers m, n, a_{ij} and b_i.

(a) $\begin{array}{rcl} x_1-2x_2+x_3 &=& 2 \\ x_2 &=& 1 \\ -x_1+x_3 &=& 1 \end{array}$ (b) $\begin{array}{rcl} x_1-3x_2 &=& 1 \\ x_2 &=& 5 \end{array}$ (c) $\begin{array}{rcl} x_1-x_2 &=& 1 \\ 2x_1-x_2 &=& 3 \\ x_2+x_1 &=& 3 \end{array}$

4. Write out the linear system that results from Example 1.1.5 if we take $n = 6$.

5. Suppose that in the input-output model of Example 1.1.6 we ignore the Materials sector input and output, so that there results a system of two equations in two unknowns x and z. Write out these equations and find a solution for them.

6. Here is an example of an economic system where everything produced by the sectors of the system is consumed by those sectors. An administrative unit has four divisions serving the internal needs of the unit, labelled (A)ccounting, (M)aintenance, (S)upplies and (T)raining. Each unit produces the “commodity” its name suggests, and charges the other divisions for its services. The fraction of commodities consumed by each division

is given by the following table , also called an “input-output matrix”.

		Produced by			
		A	M	S	T
Consumed by	A	0.2	0.1	0.4	0.4
	M	0.3	0.4	0.2	0.1
	S	0.3	0.4	0.2	0.3
	T	0.2	0.1	0.2	0.2

One wants to know what price should each division charge for its commodity so that each division earns exactly as much as it spends? Such a pricing scheme is called an *equilibrium price structure*; it assures that no division will earn too little to do its job. Let x, y, z and w be the price per unit commodity charged by A, M, S and T, respectively. The requirement of expenditures equaling earnings for each division result in a system of four equations. Find these equations.

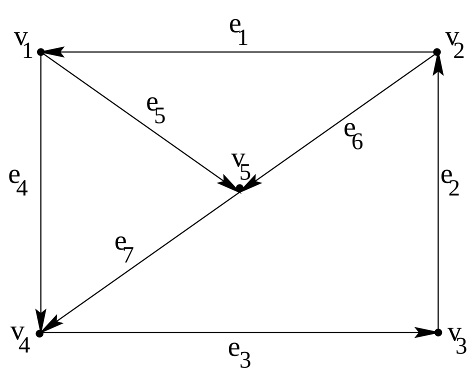
7. A polynomial $y = a + bx + cx^2$ is required to interpolate a function $f(x)$ at $x = 1, 2, 3$ where $f(1) = 1$, $f(2) = 1$ and $f(3) = 2$. Express these three conditions as a linear system of three equations in the unknowns a, b, c

8. Use your calculator, CAS or MAS to solve the system of Example 1.1.5 with known conductivity $k = 1$ and internal heat source $f(x) = x$. Then graph the approximate solution by connecting the nodes (x_j, y_j) as in Figure 1.1.3.

9. Suppose that in Example 1.1.6 the Services sector consumes all of its output. Modify the equations of the example accordingly and use your calculator, CAS or MAS to solve the system. Comment on your solution.

10. Use your calculator, CAS or MAS to solve the system of Example 1.1.6.

11. The topology of a certain network is indicated by the following graph, where five vertices (labelled v_j) represent locations of hardware units that receive and transmit data along connection edges (labelled e_j) to other units in the direction of the arrows. Suppose the system is in a steady state and that the data flow along each edge e_j is the non-negative quantity x_j. The single law that these flows must obey is this: net flow in equals net flow out at each of the five vertices (like Kirchoff’s law in electrical circuits). Write out a system of linear equations that the variables x_1, x_2, x_3, x_4, x_5 must satisfy.



1.2. Notations and a Review of Numbers

The Language of Sets

The language of sets pervades all of mathematics. It provided a convenient shorthand for expressing mathematical statements. Loosely speaking, a set can be defined as a collection of objects, called the *members* of the set. This definition will suffice for us. We use some shorthand to indicate certain relationships between sets and elements. Usually, sets will be designated by upper case letters such as A, B, etc., and elements will be designated by lower case letters such as a, b, etc. As usual, a set A is a *subset* of the set B if every element of A is an element of B, and a *proper* subset if it is a subset not equal to B. Two sets A and B are said to be *equal* if they have exactly the same elements. Some shorthand:

Set Operations

$\emptyset$ denotes the empty set, i.e., the set with no members.

$a \in A$ means "a is a member of the set A."

$A = B$ means "the set A is equal to the set B."

$A \subseteq B$ means "A is a subset of B."

$A \subset B$ means "A is a proper subset of B"

There are two ways in which we may prescribe a set: we may *list* its elements, such as in the definition $A = \{0, 1, 2, 3\}$ or specify them by *rule* such as in the definition $A = \{x \mid x \text{ is an integer and } 0 \leq x \leq 3\}$. (Read this as "$A$ is the set of x such that x is an integer and $0 \leq x \leq 3$.") With this notation we can give formal definitions of set intersections and unions:

DEFINITION 1.2.1. Let A and B be sets. Then the *intersection* of A and B is defined to be the set $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$. The *union* of A and B is the set $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$. The *difference* of A and B is the set $A - B = \{x \mid x \in A \text{ and } x \notin B\}$.

EXAMPLE 1.2.2. Let $A = \{0, 1, 3\}$ and $B = \{0, 1, 2, 4\}$. Then

$$\begin{array}{rcc} A \cup \emptyset &=& A \\ A \cap \emptyset &=& \emptyset \\ A \cup B &=& \{0, 1, 2, 3, 4\} \\ A \cap B &=& \{0, 1\} \\ A - B &=& \{3\} \end{array}$$

About Numbers

One could spend a full course fully developing the properties of number systems. We won't do that, of course, but we will review some of the basic sets of numbers, and assume the reader is familiar with properties of numbers we have not mentioned here.

At the start of it all are the kind of numbers that every child knows something about – the *natural* or *counting* numbers. This is the set

$$\mathbb{N} = \{1, 2, 3, \ldots\}$$

One could view most subsequent expansions of the concept of number as a matter of rising to the challenge of solving equations. For example, we cannot solve the equation

$$x + m = n, \ m, n \in \mathbb{N}$$

for the unknown x without introducing subtraction and extending the notion of natural number that of *integer*. The set of integers is denoted by

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}.$$

Next, we cannot solve the equation

$$ax = b, \ a, b \in \mathbb{Z}$$

for the unknown x with introducing division and extending the notion of integer to that of *rational number*. The set of rationals is denoted by

$$\mathbb{Q} = \{a/b \mid a, b \in \mathbb{Z} \text{ and} b \neq 0\}.$$

Rational number arithmetic has some characteristics that distinguish it from integer arithmetic. The main difference is that nonzero rational numbers have multiplicative inverses (the multiplicative inverse of a/b is b/a). Such a number system is called a *field* of numbers. In a nutshell, a *field of numbers* is a system of objects, called numbers, together with operations of addition, subtraction, multiplication and division that satisfy the usual arithmetic laws; in particular, it must be possible to subtract any number from any other and divide any number by a nonzero number to obtain another such number. The associative, commutative, identity and inverse laws must hold for each of addition and multiplication; and the distributive law must hold for multiplication over addition. The rationals form a field of numbers; the integers don't since division by nonzero integers is not always possible if we restrict our numbers to integers.

The jump from rational to real numbers cannot be entirely explained by algebra, although algebra offers some insight as to why the number system still needs to be extended. An equation like

$$x^2 = 2$$

does not have a rational solution, since $\sqrt{2}$ is irrational. (Story has it that this is lethal knowledge, in that followers of a Pythagorean cult claim that the gods threw overboard a ship one of their followers who was unfortunate enough to discover the fact.) There is also the problem of numbers like π and Euler's constant e which do not even satisfy any polynomial equation. The heart of the problem is that if we only consider rationals on a number line, there are many "holes" which are filled by numbers like π or $\sqrt{2}$. Filling in these holes leads us to the set $\mathbb{R}$ of real numbers, which are in one-to-one correspondence with the points on a number line. We won't give an exact definition of the set of real numbers. Recall that every real number admits a (possibly infinite) decimal representation, such as $1/3 = 0.333\ldots$ or $\pi = 3.14159\ldots$. This provides us with a loose definition: real numbers are numbers that can be expressed by a decimal representation, i.e., limits of finite decimal expansions.

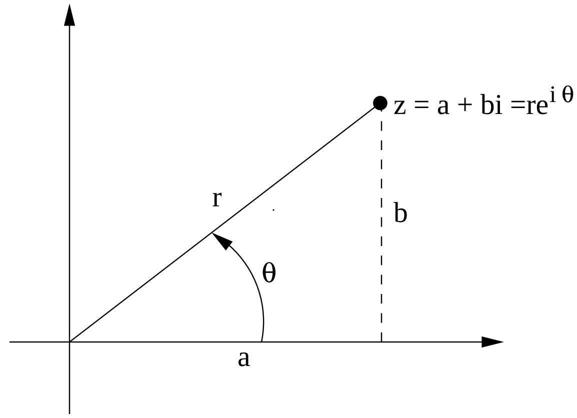


FIGURE 1.2.1. Standard and polar coordinates in the complex plane.

There is one more problem to overcome. How do we solve a system like

$$x^2 + 1 = 0$$

over the reals? The answer is we can't: if x is real, then $x^2 \geq 0$, so $x^2 + 1 > 0$. We need to extend our number system one more time, and this leads to the set $\mathbb{C}$ of *complex* numbers. We define i to be a quantity such that $i^2 = -1$ and

$$\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}.$$

Standard Form

If the complex number $z = a + bi$ is given, then we say that the form $a + bi$ is the *standard form* of z. In this case the real part of z is $\Re(z) = a$ and the imaginary part is defined as $\Im(z) = b$. (Notice that the imaginary part of z is a real number: it is the real coefficient of i.) Two complex numbers are equal precisely when they have the same real parts and the same imaginary parts. All of this could be put on a more formal basis by initially defining complex numbers to be ordered pairs of real numbers. We will not do so, but the fact that complex numbers behave like ordered pairs of real numbers leads to an important geometrical insight: complex numbers can be identified with points in the plane. Instead of an x and y axis, one lays out a *real* and *imaginary* axis (which is still usually labeled with x and y) and plots complex numbers $a + bi$ as in Figure 1.2.1. This results in the so-called *complex plane.*

Arithmetic in $\mathbb{C}$ is carried out by using the usual laws of arithmetic for $\mathbb{R}$ and the algebraic identity $i^2 = -1$ to reduce the result to standard form. Thus we have the following laws of complex arithmetic.

$$\boxed{\begin{aligned}(a + bi) + (c + di) &= (a + c) + (b + d)i, \\ (a + bi) \cdot (c + di) &= (ac - bd) + (ad + bc)i\end{aligned}}$$

In particular, notice that complex addition is exactly like the vector addition of plane vectors. Complex multiplication does not admit such a simple interpretation.

EXAMPLE 1.2.3. Let $z_1 = 2 + 4i$ and $z_2 = 1 - 3i$. Compute $z_1 - 3z_2$.

SOLUTION. We have that

$$z_1 - 3z_2 = (2+4i) - 3(1-3i) = 2+4i-3+9i = -1+13i$$

$\square$

There are several more useful ideas about complex numbers that we will need. The *length* or *absolute value* of a complex number $z = a + bi$ is defined as the nonnegative real number $|z| = \sqrt{a^2+b^2}$, which is exactly the length of z viewed as a plane vector. The *complex conjugate* of z is defined as $\overline{z} = a - bi$. Some easily checked and very useful facts about absolute value and complex conjugation:

$$\boxed{\begin{array}{ccc} |z_1 z_2| & = & |z_1|\,|z_2| \\ |z_1+z_2| & \le & |z_1|+|z_2| \\ |z|^2 & = & z\overline{z} \\ \overline{z_1+z_2} & = & \overline{z_1}+\overline{z_2} \\ \overline{z_1 z_2} & = & \overline{z_1}\,\overline{z_2} \end{array}}$$

EXAMPLE 1.2.4. Let $z_1 = 2+4i$ and $z_2 = 1-3i$. Verify for this z_1 and z_2 that $|z_1 z_2| = |z_1|\,|z_2|$.

SOLUTION. First calculate that $z_1\,z_2 = (2+4i)(1-3i) = (2+12)+(4-6)i$ so that $|z_1\,z_2| = \sqrt{14^2+(-2)^2} = \sqrt{200}$, while $|z_1| = \sqrt{2^2+4^2} = \sqrt{20}$ and $|z_2| = \sqrt{1^2+(-3)^2} = \sqrt{10}$. It follows that $|z_1 z_2| = \sqrt{10}\sqrt{20} = |z_1|\,|z_2|$. $\square$

EXAMPLE 1.2.5. Verify that the product of conjugates is the conjugate of the product.

SOLUTION. This is just the last fact in the preceding list. Let $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ be in standard form, so that $\overline{z}_1 = x_1 - iy_1$ and $\overline{z}_2 = x_2 - iy_2$. We calculate

$$z_1 z_2 = (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1)$$

so that

$$\overline{z_1 z_2} = (x_1x_2 - y_1y_2) - i(x_1y_2 + x_2y_1).$$

Also,

$$\overline{z}_1\,\overline{z}_2 = (x_1 - iy_1)(x_2 - iy_2) = (x_1x_2 - y_1y_2) + (-i(x_1y_2 - x_2y_1) = \overline{z_1 z_2}.$$

$\square$

The complex number i solves the equation $x^2+1=0$ (no surprise here: it was invented expressly for that purpose). The big surprise is that once we have the complex numbers in hand, we have a number system so complete that we can solve *any* polynomial equation in it. We won't offer a proof of this fact – it's very nontrivial. Suffice it to say that nineteenth century mathematicians considered this fact so fundamental that they dubbed it the "Fundamental Theorem of Algebra," a terminology we adopt.

Fundamental Theorem of Algebra

THEOREM 1.2.6. *Let* $p(z) = a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$ *be a non-constant polynomial in the variable* z *with complex coefficients* $a_0, \ldots, a_n$. *Then the polynomial equation* $p(z) = 0$ *has a solution in the field* $\mathbb{C}$ *of complex numbers.*

Note that the Fundamental Theorem doesn't tell us how to find a root of a polynomial – only that it can be done. As a matter of fact, there are no general formulas for the roots of a polynomial of degree greater than four, which means that we have to resort to numerical approximations in most cases.

In vector space theory the numbers in use are sometimes called *scalars*, and we will use this term. Unless otherwise stated or suggested by the presence of i, the field of scalars in which we do arithmetic is assumed to be the field of real numbers. However, we shall see later when we study eigensystems, that even if we are only interested in real scalars, complex numbers have a way of turning up quite naturally.

Let's do a few more examples of complex number manipulation.

EXAMPLE 1.2.7. Solve the linear equation $(1-2i)z = (2+4i)$ for the complex variable z. Also compute the complex conjugate and absolute value of the solution.

SOLUTION. The solution requires that we put the complex number $z = (2+4i)/(1-2i)$ in standard form. Proceed as follows: multiply both numerator and denominator by $(\overline{1-2i}) = 1+2i$ to obtain that

$$z = \frac{2+4i}{1-2i} = \frac{(2+4i)(1+2i)}{(1-2i)(1+2i)} = \frac{2-8+(4+4)i}{1+4} = \frac{-6}{5} + \frac{8}{5}i.$$

Next we see that

$$\overline{z} = \overline{\frac{-6}{5} + \frac{8}{5}i} = -\frac{6}{5} - \frac{8}{5}i$$

and

$$\begin{aligned} |z| &= \left|\frac{1}{5}(-6+8i)\right| = \frac{1}{5}\left|(-6+8i)\right| \\ &= \frac{1}{5}\sqrt{(-6)^2+8^2} = \frac{10}{5} = 2. \end{aligned}$$

□

Practical Complex Arithmetic

We conclude this section with a discussion of the more advanced aspects of complex arithmetic. This material will not be needed until Chapter 4. Recall from basic algebra the Roots Theorem: the linear polynomial $z - a$ is a factor of a polynomial $f(z) = a_0 + a_1x + \cdots a_nx^n$ if and only if a is a root of the polynomial, i.e., $f(a) = 0$. If we team this fact up with the Fundamental Theorem of Algebra, we see an interesting fact about factoring polynomials over $\mathbb{C}$: every polynomial can be completely factored into a product of linear polynomials of the form $z - a$ times a constant. The numbers a that occur are exactly the roots of $f(z)$. Of course, these roots could be repeated roots, as in the case of $f(z) = 3z^2 - 6z + 3 = 3(z-1)^2$. But how can we use the Fundamental Theorem of Algebra in a practical way to find the roots of a polynomial? Unfortunately, the usual proofs of Fundamental Theorem of Algebra don't offer a clue because they are *non-constructive*, i.e., they prove that solutions must exist, but do not show how to explicitly construct such a solution. Usually, we have to resort to numerical methods to get approximate solutions, such as the Newton's method used in calculus. For now, we will settle on a few *ad hoc* methods for solving some important special cases. First degree equations offer little difficulty: the solution to $ax = b$ is $x = b/a$, as usual. The one detail to attend to: what complex number is represented by the expression b/a ? We saw how to handle this by the trick of "rationalizing" the denominator in Example 1.2.7.

Quadratic equations are also simple enough: use the quadratic formula, which says that the solutions to Quadratic Formula

$$az^2 + bz + c = 0$$

are given by

$$z = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

There is one little catch: what does the square root of a complex number mean? What we are really asking is this: how do we solve the equation $z^2 = d$ for z, where d is a complex number? Let's try for a little more: how do we solve $z^n = d$ for all possible solutions z, where d is a given complex number? In a few cases, such an equation is quite easy to solve. We know, for example, that $z = \pm i$ are solutions to $z^2 = -1$, so these are all the solutions. Similarly, one can check by hand that $\pm 1, \pm i$ are all solutions to $z^4 = 1$. Consequently, $z^4 - 1 = (z-1)(z+1)(z-i)(z+i)$. Roots of the equation $z^n = 1$ are sometimes called the nth roots of unity. Thus the 4th roots of unity are ± 1 and $\pm i$. But what about something like $z^3 = 1 + i$?

The key to answering this question is another form of a complex number $z = a + bi$. In reference to Figure 1.1.3 we can write $z = r(\cos\theta + i\sin\theta) = re^{i\theta}$, where θ is a real number, r is a non-negative real and $e^{i\theta}$ is *defined* by the following expression: Polar form

DEFINITION 1.2.8. $e^{i\theta} \equiv \cos\theta + i\sin\theta$.

Notice that $|e^{i\theta}| = \cos^2\theta + \sin^2\theta = 1$, so that $|re^{i\theta}| = |r||e^{i\theta}| = r$, provided r is non-negative. The expression $re^{i\theta}$ with $r = |z|$ and the angle θ measured counterclockwise in radians, is called the *polar form* of z. The number r is just the absolute value of z. The number θ is sometimes called an *argument* of z. It is important to notice that θ is not unique. If the angle θ_0 works for the complex number z, then so does $\theta = \theta_0 + 2\pi k$, for any integer k, since $\sin$ and $\cos$ are periodic of period 2π. It follows that a complex number may have more than one polar form. For example, $i = e^{i\pi/2} = e^{i5\pi/2}$ (here $r = 1$). In fact, the most general polar expression for i is $i = e^{i(\pi/2 + 2k\pi)}$, where k is an arbitrary integer.

EXAMPLE 1.2.9. Find the possible polar forms of $1 + i$.

SOLUTION. Draw a picture of the number $1 + i$ as in the adjacent figure. We see that the angle $\theta_0 = \pi/4$ works fine as a measure of the angle from the positive x-axis to the radial line from the origin to z.Moreover, the absolute value of z is $\sqrt{1+1} = \sqrt{2}$. Hence, a polar form for z is $z = \sqrt{2}e^{i\pi/4}$. However, we can adjust the angle θ_0 by any multiple of 2π, a full rotation, and get a polar form for z. So the most general polar form for z is $z = \sqrt{2}e^{i(\pi/4+2k\pi)}$, where k is any integer.

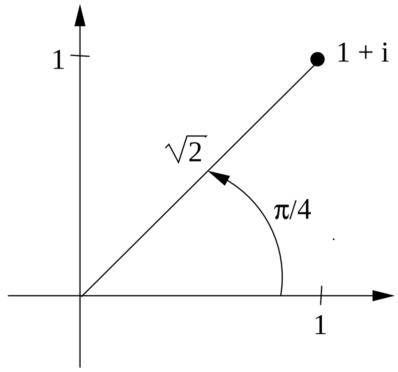


Figure 1.2.2: Form of $1 + i$

As the notation suggests, polar forms obey the laws of exponents. A simple application of the laws for the sine and cosine of a sum of angles shows that for angles θ and ψ we have the identity

$$e^{i(\theta+\psi)} = e^{i\theta}e^{i\psi}.$$

By using this formula n times, we obtain that $e^{in\theta} = (e^{i\theta})^n$ which can also be expressed as *DeMoivre's Formula:*

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$$

Now for solving $z^n = d$. First, find the general polar form of d, say $d = ae^{i(\theta_0+2k\pi)}$, where θ_0 is the so-called *principal angle* for d, i.e., $0 \leq \theta_0 < 2\pi$, and $a = |d|$. Next, write $z = re^{i\theta}$, so that the equation to be solved becomes

$$r^n e^{in\theta} = ae^{i(\theta_0+2k\pi)}.$$

Taking absolute values of both sides yields that $r^n = a$, whence we obtain the unique value of $r = \sqrt[n]{a} = \sqrt[n]{|d|}$. What about θ? The most general form for $n\theta$ is

$$n\theta = \theta_0 + 2k\pi.$$

Hence we obtain that

$$\theta = \frac{\theta_0}{n} + \frac{2k\pi}{n}.$$

Notice that the values of $e^{i2k\pi/n}$ start repeating themselves as k passes a multiple of n, since $e^{i2\pi} = e^0 = 1$. Therefore, one gets exactly n distinct values for $e^{i\theta}$, namely

$$\theta = \frac{\theta_0}{n} + \frac{2k\pi}{n}, \quad k = 0, \cdots, n-1.$$

These points are equally spaced around the unit circle in the complex plane, starting with the point $e^{i\theta_0}$. Thus we have obtained n distinct solutions to the equation $z^n = d$, where $d = ae^{i\theta_0}$, namely

General Solution to $z^n = d$

$$\boxed{z = a^{1/n}e^{i(\theta_0/n+2k\pi/n)}, \quad k = 0, \cdots, n-1}$$

EXAMPLE 1.2.10. Solve the equation $z^3 = 1 + i$ for the unknown z.

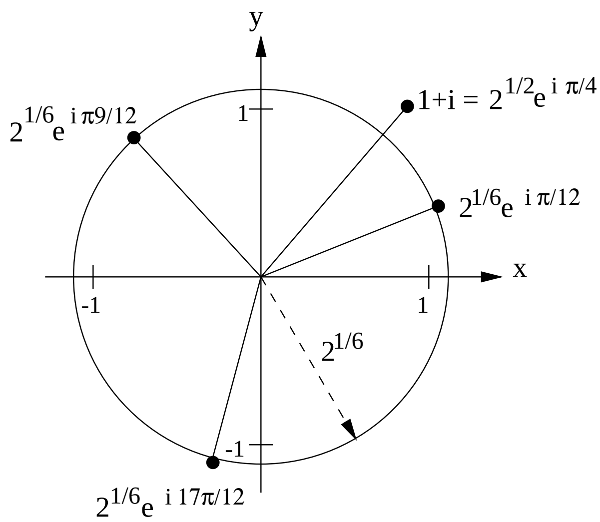


FIGURE 1.2.3. Roots of $x^3 = 1 + i$.

SOLUTION. The solution goes as follows: we have seen that $1 + i$ has a polar form

$$1 + i = \sqrt{2}e^{i\pi/4}.$$

Then according to the previous formula, the three solutions to our cubic are

$$\begin{aligned} z &= (\sqrt{2})^{1/3}e^{i(\pi/4+2k\pi)/3} \\ &= 2^{1/6}e^{i(1+8k)\pi/12} \qquad k = 0, 1, 2. \end{aligned}$$

See Figure 1.2.3 for a graph of these complex roots. □

We conclude with a little practice with square roots and the quadratic formula. In regards to square roots, notice that the expression $w = \sqrt{d}$ is ambiguous. With a positive real number d this meant the positive root of the equation $w^2 = d$. But when d is complex (or even negative), it no longer makes sense to talk about "positive" and "negative" roots of $w^2 = d$. In this case we simply interpret $\sqrt{d}$ to be one of the roots of $w^2 = d$.

EXAMPLE 1.2.11. Compute $\sqrt{-4}$ and $\sqrt{i}$.

SOLUTION. Observe that $-4 = 4 \cdot (-1)$. It is reasonable to expect the laws of exponents to continue to hold, so we should have $(-4)^{1/2} = 4^{1/2} \cdot (-1)^{1/2}$. Now we know that $i^2 = -1$, so we can take $i = (-1)^{1/2}$ and obtain that $\sqrt{-4} = (-4)^{1/2} = 2i$. Let's check it: $(2i)^2 = 4i^2 = -4$.

We have to be a bit more careful with $\sqrt{i}$. We'll just borrow the idea of the formula for solving $z^n = d$. First, put i in polar form as $i = 1 \cdot e^{i\pi/2}$. Now raise each side to the $1/2$ power to obtain

$$\begin{aligned} \sqrt{i} = i^{1/2} &= 1^{1/2} \cdot (e^{i\pi/2})^{1/2} \\ &= 1 \cdot e^{i\pi/4} = \cos(\pi/4) + i\sin(\pi/4) \\ &= \frac{1}{\sqrt{2}}(1 + i). \end{aligned}$$

A quick check confirms that $((1+i)/\sqrt{2})^2 = 2i/2 = i$. $\square$

EXAMPLE 1.2.12. Solve the equation $z^2 + z + 1 = 0$.

SOLUTION. According to the quadratic formula, the answer is

$$z = \frac{-1 \pm \sqrt{1^2 - 4}}{2} = -1 \pm i\frac{\sqrt{3}}{2}$$

$\square$

EXAMPLE 1.2.13. Solve $z^2 + z + 1 + i = 0$ and factor this polynomial.

SOLUTION. This time we obtain from the quadratic formula that

$$z = \frac{-1 \pm \sqrt{1 - 4(1+i)}}{2} = \frac{-1 \pm \sqrt{-(3+4i)}}{2}$$

What is interesting about this problem is that we don't know the polar angle θ for $z = -(3+4i)$. However, we know that $\sin\theta = -4/5$ and $\cos\theta = -3/5$. We also have the standard half angle formulas from trigonometry to help us:

$$\cos^2\theta/2 = \frac{1+\cos\theta}{2} = \frac{1}{5}, \text{and} \sin^2\theta/2 = \frac{1-\cos\theta}{2} = \frac{4}{5}$$

Since θ is in the third quadrant of the complex plane, $\theta/2$ is in the second, so

$$\cos\theta/2 = \frac{-1}{\sqrt{5}}, \text{and} \sin\theta/2 = \frac{2}{\sqrt{5}}$$

Now notice that $|-(3+4i)| = 5$. It follows that a square root of $-(3+4i)$ is given by

$$s = \sqrt{5}\left(\frac{-1}{\sqrt{5}} + \frac{2}{\sqrt{5}}i\right) = -1 + 2i$$

Check that $s^2 = -(3+4i)$. It follows that the two roots to our quadratic equation are given by

$$z = \frac{-1 \pm (-1+2i)}{2} = -1 + i, \ -i$$

In particular, we see that $z^2 + z + 1 + i = (z+1-i)(z+i)$. $\square$

1.2 Exercises

1. Given that $A = \{x | x \in \mathbb{R} \text{ and } x^2 < 3\}$ and $B = \{x | x \in \mathbb{Z} \text{ and } x > -1\}$, enumerate the following sets:

(a) $A \cap B$ (b) $B - A$ (c) $\mathbb{Z} - B$ (d) $\mathbb{N} \cup B$ (e) $\mathbb{R} \cap A$

2. Put the following complex numbers into polar form and sketch them in the complex plane:

(a) $-i$ (b) $1+i$ (c) $-1+i\sqrt{3}$ (d) -1 (e) $2-2i$ (f) $2i$ (g) π

3. Calculate the following (your answers should be in standard form):

(a) $(4+2i)-(3-6i)$ (b) $\dfrac{2+i}{2-i}$ (c) $(2+4i)(3-i)$ (d) $\dfrac{1+2i}{1-2i}$ (e) $\overline{i(1-i)}$

4. Solve the equations for the unknown z. Be sure to put your answer in standard form.

(a) $(2+i)z = 1$ (b) $-iz = 2z + 5$ (c) $\Im(z) = 2\Re(z) + 1$ (d) $\overline{z} = z$

5. Find all solutions to the equations

(a) $z^2 + z + 3 = 0$ (b) $z^2 - 1 = iz$ (c) $z^2 - 2z + i = 0$ (d) $z^2 + 4 = 0$

6. Find the solutions to the following equations. Express them in both polar and standard form and graph them in the complex plane.

(a) $z^3 = 1$ (b) $z^3 = -8$ (c) $(z-1)^3 = -1$ (d) $z^4 + z^2 + 1 = 0$

7. Write out the values of i^k in standard form for integers $k = -1, 0, 1, 2, 3, 4$ and deduce a formula for i^k consistent with these values.

8. Sketch in the complex plane the set of complex numbers z such that

(a) $|z+1| = 2$ (b) $|z+1| = |z-1|$ (c) $|z-2| < 1$

Hint: It's easier to work with absolute value squared.

9. Let $z_1 = 2 + 4i$ and $z_2 = 1 - 3i$. Verify for this z_1 and z_2 that $\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$.

10. Verify that for any two complex numbers, the sum of the conjugates is the conjugate of the sum.

11. Use the notation of Example 1.2.5 to show that $|z_1 z_2| = |z_1|\,|z_2|$. *Hint:* Remember that if $z = x + iy$ then $|z|^2 = x^2 + y^2$.

12. Use the definitions of exponentials along with the sum of angles formulas for $\sin(\theta + \psi)$ and $\cos(\theta + \psi)$ to verify the law of addition of exponents: $e^{i(\theta+\psi)} = e^{i\theta} e^{i\psi}$.

13. Use a computer or calculator to find all roots to the polynomial equation $z^5 + z + 1 = 0$. How many roots should this equation have? How many of these roots can you find with your system?

14. Show that if w is a root of the polynomial $p(z)$, that is, $p(w) = 0$, where $p(z)$ has real coefficients, then $\overline{w}$ is also a root of $p(z)$.

15. Show that $1 + i, 1 - i$ and 2 are roots of the polynomial$p(z) = z^3 - 4z^2 + 6z - 4$ and use this to factor the polynomial.

16. Show that if w is a root of the polynomial $p(z)$, that is, $p(w) = 0$, where $p(z)$ has real coefficients, then $\overline{w}$ is also a root of $p(z)$.

1.3. Gaussian Elimination: Basic Ideas

We return now to the main theme of this chapter, which is the systematic solution of linear systems, as defined in equation 1.1.1 of Section 1.1. The principal methodology is the method of *Gaussian elimination* and its variants, which we introduce by way of a few simple examples. The idea of this process is to reduce a system of equations by certain legitimate and reversible algebraic operations (called "elementary operations") to a form where we can easily see what the solutions to the system, if any, are. Specifically, we want to get the system in a form where only the first equation involves the first variable, only the first and second involve the next variable to be solved for, and so forth. Then it will be simple to solve for each variable one at a time, starting with the last equation and variable. In a nutshell, this is Gaussian elimination.

One more matter that will have an effect on our description of solutions to a linear system is that of the number system in use. As we noted earlier, it is customary in linear algebra to refer to numbers as "scalars." The two basic choices of scalar fields are the real number system or the complex number system. Unless complex numbers occur explicitly in a linear system, we will assume that the scalars to be used in finding a solution come from the field of real numbers. Such will be the case for most of the problems in this chapter.

An Example and Some Shorthand

EXAMPLE 1.3.1. Solve the simple system

$$\begin{array}{rcl} 2x - y & = & 1 \\ 4x + 4y & = & 20 \end{array} \tag{1.3.1}$$

SOLUTION. First, let's switch the equations to obtain

$$\begin{array}{rcl} 4x + 4y & = & 20 \\ 2x - y & = & 1 \end{array} \tag{1.3.2}$$

Next, multiply the first equation by $1/4$ to obtain

$$\begin{array}{rcl} x + y & = & 5 \\ 2x - y & = & 1 \end{array} \tag{1.3.3}$$

Now, multiply a copy of the first equation by -2 and add it to the second. We can do this easily if we take care to combine like terms as we go. In particular, the resulting x term in the new second equation will be $-2x + 2x = 0$, the y term will be $-2y - y = -3y$, and the constant term on the right hand side will be $-2 \cdot 5 + 1 = -9$. Thus we obtain

$$\begin{array}{rcl} x + y & = & 5 \\ 0x - 3y & = & -9 \end{array} \tag{1.3.4}$$

This completes the first phase of Gaussian elimination, which is called "forward solving." Note that we have put the system in a form where only the first equation involves the first variable and only the first and second involve the second variable. The second phase of Gaussian elimination is called "back solving", and it works like it sounds. Use the last equation to solve for the last variable, then work backwards, solving for the remaining variables in reverse order. In our case, the second equation is used to solve for y simply by dividing by -3 to obtain that

$$y = \frac{-9}{-3} = 3$$

Now that we know what y is, we can use the first equation to solve for x, and we obtain

$$x = 5 - y = 5 - 3 = 2$$

□

The preceding example may seem like too much work for such a simple system. We could easily scratch out the solution in much less space. But what if the system is larger, say 4 equations in 4 unknowns, or more? How do we proceed then? It pays to have a systematic strategy and notation. We also had an ulterior motive in the way we solved this system. All of the operations we will ever need to solve a linear system were illustrated in the preceding example: switching equations, multiplying equations by nonzero scalars, and adding a multiple of one equation to another.

Before proceeding to another example, let's work on the notation a bit. Take a closer look at the system of equations (1.3.1). As long as we write numbers down systematically, there is no need to write out all the equal signs or plus signs. Isn't every bit of information that we require contained in the following table of numbers?

$$\begin{bmatrix} 2 & -1 & 1 \\ 4 & 4 & 20 \end{bmatrix}$$

Of course, we have to remember that the first two columns of numbers are coefficients of x and y, respectively, and the third column consists of right hand side terms. So we could embellish the table with a few reminders in the top row:

$$\begin{bmatrix} x & y & = & r.h.s. \\ 2 & -1 & & 1 \\ 4 & 4 & & 20 \end{bmatrix}$$

With a little practice, we will find that the reminders are usually unnecessary; so we dispense with them for the most part. We can see that rectangular tables of numbers are very useful in representing a system of equations. Such a table is one of the basic objects studied in this text. As such, it warrants a formal definition.

DEFINITION 1.3.2. A *matrix* is a rectangular array of numbers. If a matrix has m rows and n columns, then the *size* of the matrix is said to be $m \times n$. If the matrix is $1 \times n$ or $m \times 1$, it is called a *vector*. Finally, the number that occurs in the ith row and jth column is called the $(i, j)th$ *entry* of the matrix. Matrices and Vectors

The objects we have just defined are basic "quantities" of linear algebra and matrix analysis, along with scalar quantities. Although every vector is itself a matrix, we want to single vectors out when they are identified as such. Therefore, we will follow a standard typographical convention: matrices are usually designated by capital letters, while vectors are usually designated by boldface lower case letters. In a few cases these conventions are not followed, but the meaning of the symbols should be clear from context.

We shall need to refer to parts of a matrix. As indicated above, the location of each entry of a matrix is determined by the index of the row and column it occupies.

NOTATION 1.3.3. The statement "$A = [a_{ij}]$" means that A is a matrix whose $(i, j)th$ entry is denoted by a_{ij}. Generally, the size of A will be clear from context. If we want to indicate that A is an $m \times n$ matrix, we write

$$A = [a_{ij}]_{m,n}.$$

Similarly, the statement "$\mathbf{b} = [b_i]$" means that b is a column vector whose ith entry is denoted by b_i, and "$\mathbf{c} = [c_j]$" means that c is a row vector whose jth entry is denoted by c_j. In case the type of the vector (row or column) is not clear from context, the default is a column vector.

Another term that we will use frequently is the following.

NOTATION 1.3.4. The *leading entry* of a row vector is the first nonzero element of that vector. If all entries are zero, the vector has no leading entry.

The equations of (1.3.1) have several matrices associated with them. First is the full matrix that describes the system, which we call the *augmented matrix* of the system. In our example, this is the 2×3 matrix

$$\begin{bmatrix} 2 & -1 & 1 \\ 4 & 4 & 20 \end{bmatrix}$$

Next, there is the submatrix consisting of coefficients of the variables. This is called the *coefficient matrix* of the system, and in our case it is the 2×2 matrix

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

Finally, there is the single column matrix of right hand side constants, which we call the right hand side vector. In our example, it is the 1×2 vector

$$\begin{bmatrix} 1 \\ 20 \end{bmatrix}$$

How can we describe the matrices of the general linear system of Equation 1.1.1? First, there is the $m \times n$ coefficient matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix}$$

Notice that the way we subscripted entries of this matrix is really very descriptive: the first index indicates the row position of the entry and the second, the column position of the entry. Next, there is the $m \times 1$ right hand side vector of constants

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_i \\ \vdots \\ b_m \end{bmatrix}$$

Finally, stack this matrix and vector along side each other (we use a vertical bar below to separate the two symbols) to obtain the $m \times (n+1)$ augmented matrix

$$\widetilde{A} = [A \mid \mathbf{b}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} & b_i \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} & b_m \end{bmatrix}$$

The Elementary Row Operations

There is another matter of notation that we will find extremely handy in the sequel. This is related to the operations that we performed on the preceding example. Now that we have the matrix notation we could just as well perform these operations on each row of the augmented matrix, since a row corresponds to an equation in the original system. There were three types of operations used. We shall catalogue these and give them names, so that we can document our work in solving a system of equations in a concise way. Here are the three elementary operations we shall use, described in terms of their action on rows of a matrix; an entirely equivalent description applies to the equations of the linear system whose augmented matrix is the matrix below.

Notation for Elementary Operations

- E_{ij} : This is shorthand for the elementary operation of *switching the ith and jth rows* of the matrix. For instance, in Example 1.3.1 we moved from Equation 1.3.1 to equation 1.3.2 by using the elementary operation E_{12}.
- $E_i(c)$: This is shorthand for the elementary operation of *multiplying the ith row by the nonzero constant c.* For instance, we moved from Equation 1.3.2 to (1.3.3) by using the elementary operation $E_1(1/4)$.
- $E_{ij}(d)$: This is shorthand for the elementary operation of *adding d times the jth row to the ith row.* (Read the symbols from right to left to get the right order.) For instance, we moved from Equation 1.3.3 to Equation 1.3.4 by using the elementary operation $E_{21}(-2)$.

Now let's put it all together. The whole forward solving phase of Example 1.3.1 could be described concisely with the notation we have developed:

$$\begin{bmatrix} 2 & -1 & 1 \\ 4 & 4 & 20 \end{bmatrix} \overrightarrow{E_{12}} \begin{bmatrix} 4 & 4 & 20 \\ 2 & -1 & 1 \end{bmatrix}$$

$$\overrightarrow{E_1(1/4)} \begin{bmatrix} 1 & 1 & 5 \\ 2 & -1 & 1 \end{bmatrix} \overrightarrow{E_{21}(-2)} \begin{bmatrix} 1 & 1 & 5 \\ 0 & -3 & -9 \end{bmatrix}$$

This is a big improvement over our first description of the solution. There is still the job of back solving, which is the second phase of Gaussian elimination. When doing hand calculations, we're right back to writing out a bunch of extra symbols again, which is exactly what we set out to avoid by using matrix notation.

Gauss-Jordan Elimination

Here's a better way to do the second phase by hand: stick with the augmented matrix. Starting with the last nonzero row, convert the leading entry (this means the first nonzero entry in the row) to a 1 by an elementary operation, and then use elementary operations to convert all entries above this 1 entry to 0's. Now work backwards, row by row, up to the first row. At this point we can read off the solution to the system. Let's see how it works with Example 1.3.1. Here are the details using our shorthand for elementary operations:

$$\begin{bmatrix} 1 & 1 & 5 \\ 0 & -3 & -9 \end{bmatrix} \xrightarrow[E_2(-1/3)]{} \begin{bmatrix} 1 & 1 & 5 \\ 0 & 1 & 3 \end{bmatrix} \xrightarrow[E_{12}(-1)]{} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix}$$

All we have to do is remember the function of each column in order to read off the answer from this last matrix. The underlying system that is represented is

$$\begin{aligned} 1 \cdot x + 0 \cdot y &= 2 \\ 0 \cdot x + 1 \cdot y &= 3 \end{aligned}$$

This is, of course, the answer we found earlier: $x = 2, y = 3$.

The method of combining forward and back solving into elementary operations on the augmented matrix has a name: it is called *Gauss-Jordan elimination*, and is the method of choice for solving many linear systems. Let's see how it works on an example from Section 1.1.

EXAMPLE 1.3.5. Solve the following system by Gauss-Jordan elimination.

$$\begin{aligned} x + y + z &= 4 \\ 2x + 2y + 5z &= 11 \\ 4x + 6y + 8z &= 24 \end{aligned}$$

SOLUTION. First form the augmented matrix of the system, the 3×4 matrix

$$\begin{bmatrix} 1 & 1 & 1 & 4 \\ 2 & 2 & 5 & 11 \\ 4 & 6 & 8 & 24 \end{bmatrix}$$

Now forward solve:

$$\begin{bmatrix} 1 & 1 & 1 & 4 \\ 2 & 2 & 5 & 11 \\ 4 & 6 & 8 & 24 \end{bmatrix} \xrightarrow[E_{21}(-2)]{} \begin{bmatrix} 1 & 1 & 1 & 4 \\ 0 & 0 & 3 & 3 \\ 4 & 6 & 8 & 24 \end{bmatrix}$$

$$\xrightarrow[E_{31}(-4)]{} \begin{bmatrix} 1 & 1 & 1 & 4 \\ 0 & 0 & 3 & 3 \\ 0 & 2 & 4 & 8 \end{bmatrix} \xrightarrow[E_{23}]{} \begin{bmatrix} 1 & 1 & 1 & 4 \\ 0 & 2 & 4 & 8 \\ 0 & 0 & 3 & 3 \end{bmatrix}$$

Notice, by the way, that the row switch of the third step is essential. Otherwise, we cannot use the second equation to solve for the second variable, y. Now back solve:

$$\begin{bmatrix} 1 & 1 & 1 & 4 \\ 0 & 2 & 4 & 8 \\ 0 & 0 & 3 & 3 \end{bmatrix} \xrightarrow[E_3(1/3)]{} \begin{bmatrix} 1 & 1 & 1 & 4 \\ 0 & 2 & 4 & 8 \\ 0 & 0 & 1 & 1 \end{bmatrix} \xrightarrow[E_{23}(-4)]{} \begin{bmatrix} 1 & 1 & 1 & 4 \\ 0 & 2 & 0 & 4 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\xrightarrow[E_{13}(-1)]{} \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & 2 & 0 & 4 \\ 0 & 0 & 1 & 1 \end{bmatrix} \xrightarrow[E_2(1/2)]{} \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix} \xrightarrow[E_{12}(-1)]{} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

At this point we can read off the solution to the system: $x = 1$, $y = 2$, $z = 1$. □

Systems with Non-Unique Solutions

Next, we consider an example that will pose a new kind of difficulty, namely, that of infinitely many solutions. Here is some handy terminology.

NOTATION 1.3.6. An entry of a matrix used to zero out entries above or below it by means of elementary row operations is called a *pivot*. Pivots

The entries that we use in Gaussian or Gauss-Jordan elimination for pivots are always leading entries in the row which they occupy. For the sake of emphasis, in the next few examples, we will put a circle around the pivot entries as they occur.

EXAMPLE 1.3.7. Solve for the variables x, y and z in the system

$$\begin{array}{rrrcr} x+ & y+ & z & = & 2 \\ 2x+ & 2y+ & 4z & = & 8 \\ & & z & = & 2 \end{array}$$

.

SOLUTION. Here the augmented matrix of the system is

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

Now proceed to use Gaussian elimination on the matrix.

$$\begin{bmatrix} 1 & 1 & 1 & 2 \\ 2 & 2 & 4 & 8 \\ 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow[E_{21}(-2)]{} \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

What do we do next? Neither the second nor the third row correspond to equations that involve the variable y. Switching second and third equations won't help, either. Here is the point of view that we adopt in applying Gaussian elimination to this system: the first equation has already been "used up" and is reserved for eventually solving for x. We now restrict our attention to the "unused" second and third equations. Perform the following operations to do Gauss-Jordan elimination on the system.

$$\begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow[E_2(1/2)]{} \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

$$\xrightarrow[E_{32}(-1)]{} \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[E_{12}(-1)]{} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

How do we interpret this result? We take the point of view that the first row represents an equation to be used in solving for x since the leading entry of the row is in the column of coefficients of x. By the same token, the second row represents an equation to be used in solving for z, since the leading entry of that row is in the column of coefficients of z. What about y? Notice that the third equation represented by this matrix is simply $0 = 0$, which carries no information. The point is that there is not enough information in the system to solve for the variable y, even though we started with three distinct equations. Somehow, they contained redundant information. Therefore, we take the point of view that y is not to be solved for; it is a *free* variable in the sense that it can take on any value whatsoever and yield a legitimate solution to the system. On the other hand, the variables x and z are *bound* in the sense that they will be solved for in terms of constants and free variables. The equations represented by the last matrix above are

Free and Bound Variables

$$\begin{array}{ccc} x+y & = & 0 \\ z & = & 2 \\ 0 & = & 0 \end{array}$$

Use the first equation to solve for x and the second to solve for z to obtain the *general form* of a solution to the system:

$$\begin{array}{ccr} x & = & -y \\ z & = & 2 \\ y & \text{is} & \text{free} \end{array}$$

□

In the preceding example y can take on any scalar value. For example $x = 0$, $z = 2$, $y = 0$ is a solution to the original system (check this). Likewise, $x = -5$, $z = 2$, $y = 5$ is a solution to the system. Clearly, we have an infinite number of solutions to the system, thanks to the appearance of free variables. Up to this point, the linear systems we have considered had unique solutions, so every variable was solved for, and hence bound. Another point to note, incidentally, is that the scalar field we choose to work on has an effect on our answer. The default is that y is allowed to take on any *real* value from $\mathbb{R}$. But if, for some reason, we choose to work with the complex numbers as our scalars, then y would be allowed to take on any *complex* value from $\mathbb{C}$. In this case, another solution to the system would be given by $x = -3 - i$, $z = 2$, $y = 3 + i$, for example.

To summarize, then, once we have completed Gauss-Jordan elimination on an augmented matrix, we can immediately spot the free and bound variables of the system: the column of a bound variable will have a pivot in it, while the column of a free variable will not. Another example will illustrate the point.

EXAMPLE 1.3.8. Suppose the augmented matrix of a linear system of three equations involving variables x, y, z, w becomes, after applying suitable elementary row operations,

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

Describe the general solution to the system.

SOLUTION. We solve this problem by observing that the first and third columns have pivots in them, which the second and fourth do not. The fifth column represents the right hand side. Put our little reminder labels in the matrix and we obtain

$$\left[\begin{array}{ccccc} x & y & z & w & \text{rhs} \\ 1 & 2 & 0 & -1 & 2 \\ 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Hence, x and z are bound variables, while y and w are free. The two nontrivial equations that are represented by this matrix are

$$\begin{aligned} x + 2y - w &= 2 \\ z + 3w &= 0 \end{aligned}$$

Use the first to solve for x and the second to solve for z to obtain the general solution

$$\begin{array}{ccc} x & = & 2 - 2y + w \\ z & = & -3w \\ y, w & \text{are} & \text{free} \end{array}$$

□

We have seen so far that a linear system may have exactly one solution or infinitely many. Actually, there is only one more possibility which is illustrated by the following example.

EXAMPLE 1.3.9. Solve the linear system

$$\begin{aligned} x + y &= 1 \\ 2x + y &= 2 \\ 3x + 2y &= 5 \end{aligned}$$

We extract the augmented matrix and proceed with Gauss-Jordan elimination. This time we'll save a little space by writing more than one elementary operation between matrices. It is understood that they are done in order, starting with the top one. This is a very efficient way of doing hand calculations and minimizing the amount of rewriting of matrices as we go.

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 2 \\ 3 & 2 & 5 \end{bmatrix} \xrightarrow[E_{31}(-3)]{E_{21}(-2)} \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & -1 & 2 \end{bmatrix} \xrightarrow[E_{32}(-1)]{} \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Stop everything! We aren't done with Gauss-Jordan elimination yet since we've only done the forward solving portion. But something strange is going on here. Notice that the third row of the last matrix above stands for the equation $0x + 0y = 2$, i.e., $0 = 2$. This is impossible. What this matrix is telling us is that the original system has no solution, i.e., it is *inconsistent*. A system can be identified as inconsistent as soon as one encounters a leading entry in the column of constant terms. For this always means that an equation of the form $0 =$nonzero constant has been formed from the system by legitimate algebraic operations. Thus, one needs proceed no further. The system has no solutions. □

DEFINITION 1.3.10. A system of equations is *consistent* if it has at least one solution. Otherwise it is called *inconsistent*. Consistent Systems

Our last example is one involving complex numbers explicitly.

EXAMPLE 1.3.11. Solve the following system of equations:

$$\begin{aligned} x+y &= 4 \\ (-1+i)x+y &= -1 \end{aligned}$$

SOLUTION. The procedure is the same, no matter what the field of scalars is. Of course, the arithmetic is a bit harder. Gauss-Jordan elimination yields

$$\begin{bmatrix} 1 & 1 & 4 \\ -1+i & 1 & -1 \end{bmatrix} \xrightarrow[E_{21}(1-i)]{} \begin{bmatrix} 1 & 1 & 4 \\ 0 & 2-i & 3-4i \end{bmatrix}$$

$$\xrightarrow[E_2(1/(2-i))]{} \begin{bmatrix} 1 & 1 & 4 \\ 0 & 1 & 2-i \end{bmatrix} \xrightarrow[E_{12}(-1)]{} \begin{bmatrix} 1 & 0 & 2+i \\ 0 & 1 & 2-i \end{bmatrix}$$

Here we used the fact that

$$\frac{3-4i}{2-i} = \frac{(3-4i)(2+i)}{(2-i)(2+i)} = \frac{10-5i}{5} = 2-i$$

Thus, we see that the system has unique solution

$$\begin{aligned} x &= 2+i \\ y &= 2-i \end{aligned}$$

$\square$

1.3 Exercises

1. For each of the following matrices identify the size and the (i, j)th entry for all relevant indices i and j.

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

2. Exhibit the augmented matrix of each system and give its size. Then use Gaussian elimination and backsolving to find the general solution to the systems.

(a) $\begin{aligned} 2x+3y &= 7 \\ x+2y &= -2 \end{aligned}$ (b) $\begin{aligned} 3x_1+6x_2-x_3 &= -4 \\ -2x_1-4x_2+x_3 &= 3 \\ x_3 &= 1 \end{aligned}$ (c) $\begin{aligned} x_1+x_2 &= -2 \\ 5x_1+2x_2 &= 5 \\ x_1+2x_2 &= -7 \end{aligned}$

3. Exhibit the augmented matrix of each system and give its size. Then use Gauss-Jordan elimination to find the general solution to the systems.

(a) $\begin{aligned} x_1+x_2+x_4 &= 1 \\ 2x_1+2x_2+x_3+x_4 &= 1 \\ 2x_1+2x_2+2x_4 &= 2 \end{aligned}$ (b) $\begin{aligned} x_3+x_4 &= 0 \\ -2x_1-4x_2+x_3 &= 0 \\ 3x_1+6x_2-x_3+x_4 &= 0 \end{aligned}$ (c) $\begin{aligned} x_1+x_2+3x_3 &= 2 \\ 2x_1+5x_2+9x_3 &= 1 \\ x_1+2x_2+4x_3 &= 1 \end{aligned}$

4. Each of the following matrices results from applying Gauss-Jordan elimination to the augmented matrix of a linear system. In each case, write out the general solution to the system or indicate that it is inconsistent.

$$\text{(a)} \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 2 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

5. Use any method to find the solution to each of the following systems. Here, b_1, b_2 are constants and x_1, x_2 are the unknowns.

$$\text{(a)} \begin{aligned} x_1 - x_2 &= b_1 \\ x_1 + 2x_2 &= b_2 \end{aligned} \quad \text{(b)} \begin{aligned} x_1 - x_2 &= b_1 \\ 2x_1 - 2x_2 &= b_2 \end{aligned} \quad \text{(c)} \begin{aligned} ix_1 - x_2 &= b_1 \\ 2x_1 + 2x_2 &= b_2 \end{aligned}$$

6. Use Gauss-Jordan elimination to find the general solution to these systems. Show the elementary operations you use.

$$\text{(a)} \begin{aligned} 2x_1 + x_2 + 7x_3 &= -1 \\ 3x_1 + 2x_2 - 2x_4 &= 1 \\ 2x_1 + 2x_2 + 2x_3 - 2x_4 &= 4 \end{aligned} \quad \text{(b)} \begin{aligned} x_1 + x_2 + x_3 - x_4 &= 2 \\ 2x_1 + x_2 - 2x_4 &= 1 \\ 2x_1 + 2x_2 + 2x_3 - 2x_4 &= 4 \end{aligned}$$

7. Exercise 6 of Section 1.1 led to the following system. Solve it and see if there exists a nontrivial solution consisting of *positive* numbers. Why is this important for the problem?

$$\begin{aligned} 8x - y - 4z - 4w &= 0 \\ -3x + 6y - 2z - 1w &= 0 \\ -3x - 4y + 8z - 3w &= 0 \\ -2x - 1y - 2z + 8w &= 0 \end{aligned}$$

8. Apply the operations found in Exercise 6 in the same order to right hand side vector $b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ .What does this tell you about the system's consistency?

9. Suppose that we want to solve the three systems with the same left hand side

$$\text{(a)} \begin{aligned} x_1 + x_2 &= 1 \\ x_2 + 2x_3 &= 0 \\ 2x_2 + x_3 &= 0 \end{aligned} \quad \text{(b)} \begin{aligned} x_1 + x_2 &= 0 \\ x_2 + 2x_3 &= 0 \\ 2x_2 + x_3 &= 0 \end{aligned} \quad \text{(c)} \begin{aligned} x_1 + x_2 &= 2 \\ x_2 + 2x_3 &= 3 \\ 2x_2 + x_3 &= 3 \end{aligned}$$

Show how to do this efficiently by using only one augmented matrix consisting of the common coefficient matrix and the three right hand sides stacked along side each other.

10. Show that the following nonlinear systems become linear if we view the unknowns as $1/x, 1/y$ and $1/z$ rather than x, y and z. Use this to find the solution sets of the nonlinear systems. (You must also account for the possibilities that one of x, y, z is zero.)

$$\text{(a)} \begin{aligned} 2x - y + 3xy &= 0 \\ 4x + 2y - xy &= 0 \end{aligned} \qquad \text{(b)} \begin{aligned} yz + 3xz - xy &= 0 \\ yz + 2xy &= 0 \end{aligned}$$

11. Use a computer program or calculator with linear algebra capabilities (such as Derive, Maple, Mathematica, Macsyma, MATLAB, TI-85, HP48, etc.) to solve the system of Example 1.1.5 with $n = 8$ and $f(x) = x$.

12. Write out the system derived from the input-output model of page 7 and use your computer system or calculator to solve it. Is the solution physically meaningful?

13. Solve the linear system that was found in Exercise 11 on page 8. Does this data network have any steady state solutions?

14. Suppose the function $f(x)$ is to be interpolated at three interpolating points x_0, x_1, x_2 by a quadratic polynomial $p(x) = a + bx + cx^2$, that is, $f(x_i) = p(x_i), i = 0, 1, 2$. As in Exercise 7 of Section 1.1, this leads to a system of three linear equations in the three unknowns a, b, c.

(a) Write out these equations.

(b) Apply the equations of part (a) to the specific $f(x) = \sin(x), 0 \leq x \leq \pi$ with x_j equal $0, \pi/2, \pi$, and graph the resulting quadratic against $f(x)$.

(c) Plot the error function $f(x)-p(x)$ and estimate the largest value of the error function by trial and error.

(d) Find three points x_1, x_2, x_3 on the interval $0 \leq x \leq \pi$ for which the resulting interpolating quadratic gives an error function with a smaller largest value than that found in part (c).

15. Solve the network system of Exercise 11 and exhibit all physically meaningful solutions.

1.4. Gaussian Elimination: General Procedure

The preceding section introduced Gaussian elimination and Gauss-Jordan elimination at a practical level. In this section we will see why these methods work and what they really mean in matrix terms. Then we will find conditions of a very general nature under which a linear system has (none, one or infinitely many) solutions. A key idea that comes out of this section is the notion of the *rank* of a matrix.

Equivalent Systems

The first question to be considered is this: how is it that Gaussian elimination or Gauss-Jordan elimination gives us *every* solution of the system we begin with and *only* solutions to that system? To see that linear systems are special, consider the following nonlinear system of equations.

EXAMPLE 1.4.1. Solve for the real roots of the system

$$\begin{aligned} x + y &= 2 \\ \sqrt{x} &= y \end{aligned}$$

SOLUTION. Let's follow the Gauss-Jordan elimination philosophy of using one equation to solve for one unknown. So the first equation enables us to solve for y to get $y = 2 - x$. Next substitute this into the second equation to obtain $\sqrt{x} = 2 - x$. Then square both sides to obtain $x = (2 - x)^2$, or

$$0 = x^2 - 5x - 4 = (x - 1)(x - 4)$$

Now $x = 1$ leads to $y = 1$, which is a solution to the system. But $x = 4$ gives $y = -2$, which is not a solution to the system since $\sqrt{x}$ cannot be negative. □

What went wrong in this example is that the squaring step introduced extraneous solutions to the system. Why is Gaussian or Gauss-Jordan elimination safe from this kind of difficulty? The answer lies in examining the kinds of operations we perform with these methods. First, we need some terminology. Up to this point we have always described a solution to a linear system in terms of a list of equations. For general problems this is a bit of a nuisance. Since we are using the matrix/vector notation, we may as well go all the way and use it to concisely describe solutions as well. We will use column vectors to define solutions as follows.

DEFINITION 1.4.2. A *solution vector* for the general linear system given by Equation 1.1.1 is a vector

$$\mathbf{x} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}$$

such that the resulting equations are satisfied for these choices of the variables. The set of all such solutions is called the *solution set* of the linear system, and two linear systems are said to be *equivalent* if they have the same solution sets.

We will want to make frequent reference to vectors without having to display them in the text. Of course, for row vectors ($1 \times n$) this is no problem. To save space in referring to column vectors, we shall adopt the convention that a column vector will also be denoted by a tuple with the same entries.

NOTATION 1.4.3. The n-tuple $(x_1, x_2, \dots, x_n)$ is a shorthand for the $n \times 1$ column vector $\mathbf{x}$ with entries $x_1, x_2, \dots, x_n$.

For example, we can write $(1, 3, 2)$ in place of

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

EXAMPLE 1.4.4. Describe the solution sets of all the examples worked out in the previous section.

SOLUTION. Here is the solution set to Example 1.3.1. It is the singleton set

$$S = \left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\} = \{(2, 3)\}$$

The solution set for Example 1.3.5 is $S = \{(1, 2, 1)\}$. (Remember that we can designate column vectors by tuples if we wish.)

For Example 1.3.7 the solution set requires some fancier set notation, since it is an infinite set. Here it is :

$$S = \left\{ \begin{bmatrix} -y \\ y \\ 2 \end{bmatrix} \mid y \in \mathbb{R} \right\} = \{(-y, y, 2) \mid y \in \mathbb{R}\}$$

Example 1.3.9 was an inconsistent system, so had no solutions. Hence its solution set is $S = \emptyset$.

Finally, the solution set for Example 1.3.11 is the singleton set $S = \{(2+i, 2-i)\}$. $\square$

A key question about Gaussian elimination and equivalent systems: what happens to a system if we change it by performing one elementary row operation? After all, Gaussian and Gauss-Jordan elimination amount to a sequence of elementary row operations applied to the augmented matrix of a given linear system. The answer: nothing happens to the solution set!

THEOREM 1.4.5. *Suppose a linear system has augmented matrix A upon which an elementary row operation is applied to yield a new augmented matrix B corresponding to a new linear system. Then these two linear systems are equivalent, i.e., have the same solution set.*

PROOF. If we replace the variables in the system corresponding to A by the values of a solution, the resulting equations will be satisfied. Now perform the elementary operation in question on this system of equations to obtain that the equations for the system corresponding to the augmented matrix B are also satisfied. Thus, every solution to the old system is also a solution to the new system resulting from performing an elementary operation. It is sufficient for us to show that the old system can be obtained from the new one by another elementary operation. In other words, we need to show that the effect of any elementary operation can be undone by another elementary operation. This will show that every solution to the new system is also a solution to the old system. If E represents an elementary operation, then the operation that undoes it could reasonably be designated as E^{-1}, since the effect of the inverse operation is rather like cancelling a number by multiplying by its inverse. Let us examine each elementary operation in turn.

Inverse Elementary Operations

- E_{ij} : The elementary operation of switching the ith and jth rows of the matrix. Notice that the effect of this operation is undone by performing the same operation, E_{ij}, again. This switches the rows back. Symbolically we write $E_{ij}^{-1} = E_{ij}$.
- $E_i(c)$: The elementary operation of multiplying the ith row by the nonzero constant c. This elementary operation is undone by performing the elementary operation $E_i(1/c)$; in other words, by dividing the ith row by the nonzero constant c. We write $E_i(c)^{-1} = E_i(1/c)$.
- $E_{ij}(d)$: The elementary operation of adding d times the jth row to the ith row. This operation is undone by subtracting d times the jth row to the ith row. We write $E_{ij}(d)^{-1} = E_{ij}(-d)$.

Thus, in all cases the effects of an elementary operation can be undone by applying another elementary operation of the same type, which is what we wanted to show. $\square$

The inverse notation we used here doesn't do much for us yet. In Chapter 2 this notation will take on an entirely new and richer meaning.

The Reduced Row Echelon Form

Theorem 1.4.5 tells us that the methods of Gaussian or Gauss-Jordan elimination do not alter the solution set we are interested in finding. Our next objective is to describe the end result of these methods in a precise way. That is, we want to give a careful definition of the form of the matrix that these methods lead us to, starting with the augmented matrix of the original system. Recall that the *leading entry* of a row is the first nonzero entry of that row. (So a row of zeros has no leading entry.)

DEFINITION 1.4.6. A matrix R is said to be in *reduced row form* if:

(1) The nonzero rows of R precede the zero rows.
(2) The column numbers of the leading entries of the nonzero rows, say rows $1, 2, ..., r$, form an increasing sequence of numbers $c_1 < c_2 < \cdots < c_r$.

The matrix R said to be in *reduced row echelon form* if, in addition to the above:

(3) Each leading entry is a 1.
(4) Each leading entry has only zeros above it.

EXAMPLE 1.4.7. Consider the following matrices (whose leading entries are enclosed in a circle). Which are in reduced row form? reduced row echelon form?

$$\text{(a)} \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \qquad \text{(c)} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\text{(d)} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{(e)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

SOLUTION. Checking through (1)-(2), we see that (a), (b) and (d) fulfill all the conditions for reduced row matrices. But (c) fails, since a zero row precedes the nonzero ones; matrix (e) fails to be reduced row form because the column numbers of the leading entries do not form an increasing sequence. Matrices (a) and (b) don't satisfy (3), so matrix (d) is the only one that satisfies (3)-(4). Hence, it is the only matrix in the list in reduced row echelon form. □

We can now describe the goal of Gaussian elimination as follows: use elementary row operations to reduce the augmented matrix of a linear system to reduced row form; then back solve the resulting system. On the other hand, the goal of Gauss-Jordan elimination is to use elementary operations to reduce the augmented matrix of a linear system to reduced row echelon form. From this form one can read off the solution(s) to the system.

Is it always possible to reduce a matrix to a reduced row form or row echelon form? If so, how many? These are important questions because, when we take the matrix in question to be the augmented matrix of a linear system, what we are really asking becomes: does Gaussian elimination always work on a linear system? If so, do they lead us to answers that have the same form? Notice how the last question was phrased. We know that the solution set of a linear system is unaffected by elementary row operations. Therefore, the solution sets we obtain will always be the same with either method, *as sets*. But couldn't the form change? For instance, in Example 1.3.7 we obtained a form for the general solution that involved one free variable, y, and two bound variables x and z. Is it possible that by a different sequence of elementary operations we could have reduced to a form where there were two free variables and only one bound variable? This would be a rather different form, even though it might lead to the same solution set.

Certainly, matrices can be transformed by elementary row operations to different reduced row forms, as the following simple example shows:

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 2 & -1 \end{bmatrix} \xrightarrow[E_{12}(-1)]{} \begin{bmatrix} 1 & 0 & 5 \\ 0 & 2 & -1 \end{bmatrix} \xrightarrow[E_2(1/2)]{} \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & -1/2 \end{bmatrix}$$

Every matrix of this example is already in reduced row form. The last matrix is also in reduced row echelon form. Yet all three of these matrices can be obtained from each other by elementary row operations. It is significant that only one of the three matrices is in reduced row echelon form. As a matter of fact, any matrix can be reduced by elementary row operations to *one and only one* reduced row echelon form, which we can call *the* reduced row echelon form of the given matrix. The example above shows that the matrix A has as its reduced row echelon form the matrix $E = \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & -1/2 \end{bmatrix}$. Our assertions are justified by the following fundamental theorem about matrices.

Uniqueness of Reduced Row Echelon Form

THEOREM 1.4.8. *Every matrix can be reduced by a sequence of elementary row operations to one and only one reduced row echelon form.*

PROOF. First we show that every $m \times n$ matrix A can be reduced to some reduced row echelon form. Here is the algorithm we have been using: given that the first s columns of A are in reduced row echelon form with r nonzero rows and that $r < m$ and $s < n$, find the smallest column number j such that $a_{ij} \neq 0$ and $i > r, j > s$. If none is found, A is in reduced row echelon form. Otherwise, interchange rows i and $r+1$, then use elementary row operations to convert $a_{r+1,j}$ to 1, and to zero out the entries above and below this one. Now set $s = j$ and increment r by one. Continue this procedure until $r = m$ or $s = n$. This must occur at some point since both r and s increase with each step, and when it occurs, the resulting matrix is in reduced row echelon form.

Next, we prove uniqueness. Suppose that some matrix could be reduced to two distinct reduced row echelon forms. We show this is impossible. If it were possible, we could find an example $m \times n$ matrix $\widetilde{A}$ with the fewest possible columns n; that is, the theorem is true for every matrix with fewer columns. Then $n > 1$, since a single column matrix can be reduced to only one reduced row echelon form, namely either the 0 column or a column with first entry 1 and the other entries 0. Now $\widetilde{A}$ can be reduced to two reduced row echelon forms, say R_1 and R_2, with $R_1 \neq R_2$. Write $\widetilde{A} = [A \mid \mathbf{b}]$ so that we can think of $\widetilde{A}$ as the augmented matrix of a linear system (1.1.1). Now for $i = 1, 2$ write each R_i as $R_i = [L_i \mid \mathbf{b}_i]$, where $\mathbf{b}_i$ is the last column of the $m \times n$ matrix R_i, and L_i is the $m \times (n-1)$ matrix formed from the first $n-1$ columns of R_i. Each L_i satisfies the definition of reduced row echelon form, since each R_i is in reduced row echelon form. Also, each L_i results from performing elementary row operations on the matrix

A, which has only $n-1$ columns. By the minimum columns hypothesis, we have that $L_1 = L_2$. There are two possibilities to consider.

Case 1: The last column b_i of either R_i has a leading entry in it. Then the system of equations represented by $\widetilde{A}$ is inconsistent. It follows that both columns $\mathbf{b}_i$ have a leading entry in them, which must be a 1 in the first row whose portion in L_i consists of zeros, and the entries above and below this leading entry must be 0. Since $L_1 = L_2$, it follows that $\mathbf{b}_1 = \mathbf{b}_2$, and thus $R_1 = R_2$, a contradiction. So this case can't occur.

Case 2: Each b_i has no leading entry in it. Then the system of equations represented by $\widetilde{A}$ is consistent. Both augmented matrices have the same basic and free variables since $L_1 = L_2$. Hence we obtain the same solution with either augmented matrix by setting the free variables of the system equal to 0. When we do so, the bound variables are uniquely determined: the first equation says that the first bound variable equals the first entry in the right hand side vector, the second says that the second bound variable equals the second entry in the right hand side vector, and so forth. Whether we use R_1 or R_2 to solve the system, we obtain the same result, since we can manipulate one such solution into the other by elementary row operations. Therefore, $\mathbf{b}_1 = \mathbf{b}_2$ and thus $R_1 = R_2$, a contradiction again. Hence, there can be no counterexample to the theorem, which completes the proof. $\square$

The following consequence of the preceding theorem is a fact that we will find helpful in Chapter 2.

COROLLARY 1.4.9. *Let the matrix B be obtained from the matrix A by performing a sequence of elementary row operations on A. Then B and A have the same reduced row echelon form.*

PROOF. We can obtain the reduced row echelon form of B in the following manner: first perform the elementary operations on B that undo the ones originally performed on A to get B. The matrix A results from these operations. Now perform whatever elementary row operations are needed to reduce A to its reduced row echelon form. Since B can be reduced to one and only one reduced row echelon form, the reduced row echelon forms of A and B coincide, which is what we wanted to show. $\square$

Rank and Nullity of a Matrix

Now that we have Theorem 1.4.8 in hand, we can introduce the notion of *rank* of a matrix. Since A can be reduced to one and only one reduced row echelon form by Theorem 1.4.8, we see that the following definition is unambiguous.

DEFINITION 1.4.10. The *rank* of a matrix A is the number of nonzero rows of the reduced row echelon form of A. This number is written as $\operatorname{rank} A$.

Rank can also be defined as the number of nonzero rows in any reduced row form of a matrix. One has to check that any two reduced row forms have the same number of nonzero rows. Notice that the rank can also be defined as the number of columns of the reduced row echelon form with leading entries in them, since each leading entry of a reduced row echelon form occupies a unique column. We can count up the other columns as well.

DEFINITION 1.4.11. The *nullity* of a matrix A is the number of columns of the reduced row echelon form of A that do not contain a leading entry. This number is written as $\operatorname{null} A$.

In the case that A is the coefficient matrix of a linear system, we can interpret the rank of A as the number of bound variables of the system and the nullity of A as the number of free variables of the system.

Observe that the rank of a matrix is a non-negative number. But it *could* be 0! This happens when the matrix is a zero matrix, so that it has no nonzero rows. In this case, the nullity of the matrix is as large as possible. Here are some simple limits on the size of $\operatorname{rank} A$ and $\operatorname{null} A$. In one limit we shall use a notation that occurs frequently throughout the text, so we explain it first.

NOTATION 1.4.12. For a list of real numbers $a_1, a_2, \ldots, a_m$, $\min\{a_1, a_2, \ldots, a_m\}$ means the smallest number in the list and $\max\{a_1, a_2, \ldots, a_m\}$ means the largest number in the list.

THEOREM 1.4.13. *Let A be an $m \times n$ matrix. Then*

1. $0 \le \operatorname{rank} A \le \min\{m, n\}$.
2. $\operatorname{rank} A + \operatorname{null} A = n$.

PROOF. By definition, $\operatorname{rank} A$ is the number of nonzero rows of the reduced row echelon form of A, which is itself an $m \times n$ matrix. There can be no more leading entries than rows, hence $\operatorname{rank} A \le m$. Also, each leading entry of a matrix in reduced row echelon form is the unique nonzero entry in its column. Therefore, there can be no more leading entries than columns n. Since $\operatorname{rank} A$ is less than or equal to both m and n, it must be less than or equal to their minimum, which is what the first inequality says. Also notice that every column of A either has a pivot in it or not. The number of pivot columns is just $\operatorname{rank} A$ and the number of non-pivot columns is $\operatorname{null} A$. Hence the sum of these numbers is n. □

Full Column Rank

In words, part 1 of Theorem 1.4.13 says that the rank of a matrix cannot exceed the number of rows or columns of the matrix. One situation occurs often enough enough that it is entitled to its own name: if the rank of a matrix equals its column number we say that the matrix has full column rank. One has to be a little careful about this idea of rank. Consider the following example.

EXAMPLE 1.4.14. Find the rank and nullity of the matrix

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 5 \\ 3 & 3 & 7 \end{bmatrix}$$

SOLUTION. We know that the rank is at most 3 by the preceding theorem. Elementary row operations give

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 5 \\ 3 & 3 & 7 \end{bmatrix} \xrightarrow[E_{21}(-2)]{} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 3 & 3 & 2 \end{bmatrix} \xrightarrow[E_{31}(-3)]{} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & -4 \end{bmatrix}$$

$$\xrightarrow[E_{12}(-2)]{E_{32}(4)} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

From the reduced row echelon form of A at the far right we see that the rank of A is 2. □

Notice that one can't simply count the number of nonzero rows of A, which in this case is 3, to get the rank of A.

Caution: Remember that the rank of A is the number of nonzero rows of its reduced row form, and *not* the number of nonzero rows of A itself.

The notion of rank of a matrix gives us some more leverage on the question of how the solution set of a linear system behaves.

Consistency in Terms of Rank

THEOREM 1.4.15. *The general linear system 1.1.1 with $m \times n$ coefficient matrix A, right hand side vector $\mathbf{b}$ and augmented matrix $\widetilde{A} = [A \mid \mathbf{b}]$ is consistent if and only if $\operatorname{rank} A = \operatorname{rank} \widetilde{A}$, in which case either*

1. $\operatorname{rank} A = n$, *in which case the system has a unique solution, or*
2. $\operatorname{rank} A < n$, *in which case the system has infinitely many solutions.*

PROOF. We can reduce $\widetilde{A}$ to reduced row echelon form by first doing the elementary operations that reduce the A part of the matrix to reduced row echelon form, then attending to the last column. Hence, it is always the case that $\operatorname{rank} A \leq \operatorname{rank} \widetilde{A}$. The only way to get strict inequality is to have a leading entry in the last column, which means that some equation in the equivalent system corresponding to the reduced augmented matrix is $0 = 1$, which implies that the system is inconsistent. On the other hand, we have already seen (in the proof of Theorem 1.4.8, for example) that if the last column does not contain a leading entry, then the system is consistent. This establishes the first statement of the theorem.

Now suppose that $\operatorname{rank} A = \operatorname{rank} \widetilde{A}$, so that the system is consistent. By Theorem 1.4.13, $\operatorname{rank} A \leq n$, so that either $\operatorname{rank} A < n$ or $\operatorname{rank} A = n$. The number of variables of the system is n. Also, the number of leading entries (equivalently, pivots) of the reduced row form of $\widetilde{A}$, which is $\operatorname{rank} A$, is equal to the number of bound variables; the remaining $n - \operatorname{rank} A$ variables are the free variables of the system. Thus, to say that $\operatorname{rank} A = n$, is to say that no variables are free; that is, solving the system leads to a unique solution. And to say that $\operatorname{rank} A < n$ is to say that there is at least one free variable in which case the system has infinitely many solutions. □

Here is an example of how this theorem can be put to work. It confirms our intuition that if a system does not have "enough" equations, then it can't have a unique solution.

COROLLARY 1.4.16. *If a consistent linear system of equations has more unknowns than equations, then the system has infinitely many solutions.*

PROOF. In the notation of the previous theorem, the hypothesis simply means that $m < n$. But we know from Theorem 1.4.13 that $\operatorname{rank} A \leq \min\{m, n\}$. Thus $\operatorname{rank} A < n$ and the last part of Theorem 1.4.15 applies to give the desired result. □

Of course, there is still the question of when a system is consistent. In general, there isn't an easy way to see when this is so. However, in special cases we can answer the question easily. One such important special case is given by the following definition.

Homogeneous Systems

DEFINITION 1.4.17. The general linear system 1.1.1 with $m \times n$ coefficient matrix A and right hand side vector $\mathbf{b}$ is said to be *homogeneous* if the entries of $\mathbf{b}$ are all zero. Otherwise, the system is said to be *non-homogeneous*.

The nice feature of homogeneous systems is that they are always consistent! In fact, it is easy to exhibit a specific solution to the system, namely, take the value of all the variables to be zero. For obvious reasons this solution is called the *trivial* solution to the system. Thus, the previous corollary implies that a homogeneous linear system with fewer equations than unknowns must have infinitely many solutions. Of course, if we want to find all the solutions, we will have to do the work of Gauss-Jordan elimination. However, we acquire a small notational convenience in dealing with homogeneous systems. Notice that the right hand side of zeros is never changed by an elementary row operation. So why bother writing out the augmented matrix of such a system? It suffices to perform elementary operations on the coefficient matrix alone. In the end, the right hand side is still a column of zeros.

EXAMPLE 1.4.18. Solve and describe the solution set of the homogeneous system

$$\begin{aligned} x_1 + x_2 + x_4 &= 0 \\ x_1 + x_2 + 2x_3 &= 0 \\ x_1 + x_2 \qquad &= 0 \end{aligned}$$

SOLUTION. In this case we only perform row operations on the coefficient matrix to obtain

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \xrightarrow[E_{31}(-1)]{E_{21}(-1)} \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & -1 \end{bmatrix} \xrightarrow[E_{3}(-1)]{E_{2}(1/2)} \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1/2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow[E_{13}(-1)]{E_{23}(1/2)} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

One has to be a little careful here: the leading entry in the last column does not indicate that the system is inconsistent. Had we carried the right hand side column along in the calculations above, we would have obtained

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

which is the matrix of a consistent system. We see from the reduced row echelon form of the coefficient matrix that x_2 is free and the other variables are bound. The general solution is

$$\begin{aligned} x_1 &= -x_2 \\ x_3 &= 0 \\ x_4 &= 0 \\ x_2 &\text{ is free.} \end{aligned}$$

Finally, the solution set S of this system can be described as

$$S = \{(-x_2, 0, 0, x_2) \mid x_2 \in \mathbb{R}\}.$$

□

1.4 Exercises

1. Circle leading entries and determine which of the following matrices are in reduced row echelon form.

$$\text{(a)} \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 2 \end{bmatrix}$$

$$\text{(d)} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{(e)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{(f)} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

2. Show the elementary operations you use to find the reduced row echelon form of the following matrices. Give the rank and nullity of each matrix.

$$\text{(a)} \begin{bmatrix} 1 & -1 & 2 \\ 1 & 3 & 4 \\ 2 & 2 & 6 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 3 & 1 & 9 & 2 \\ -3 & 0 & 6 & -5 \\ 0 & 0 & 1 & 2 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} 0 & 1 & 0 & 1 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$\text{(d)} \begin{bmatrix} 2 & 4 & 2 \\ 4 & 9 & 3 \\ 2 & 3 & 3 \end{bmatrix} \quad \text{(e)} \begin{bmatrix} 2 & 2 & 5 & 6 \\ 1 & 1 & -2 & 2 \end{bmatrix} \quad \text{(f)} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

3. Find the rank of the augmented and coefficient matrix of the following linear systems and the solution to to following systems.

$$\text{(a)} \quad \begin{aligned} x_1 + x_2 + x_3 - x_4 &= 2 \\ 2x_1 + x_2 - 2x_4 &= 1 \\ 2x_1 + 2x_2 + 2x_3 - 2x_4 &= 4 \end{aligned} \qquad \text{(b)} \quad \begin{aligned} x_3 + x_4 &= 0 \\ -2x_1 - 4x_2 &= 0 \\ 3x_1 + 6x_2 - x_3 + x_4 &= 0 \end{aligned}$$

4. Consider two systems of equations

$$\text{(A)} \quad \begin{aligned} x_1 + x_2 + x_3 - x_4 &= 2 \\ 2x_1 + x_2 - 2x_4 &= 1 \\ 2x_1 + 2x_2 + 2x_3 - 2x_4 &= 4 \end{aligned} \qquad \text{(B)} \quad \begin{aligned} x_3 + x_4 &= 0 \\ -2x_1 - 4x_2 &= 0 \\ 3x_1 + 6x_2 - x_3 + x_4 &= 0 \end{aligned}$$

(a) Find a sequence of elementary operations that transforms system (A) into (B).

(b) It follows that these two systems are equivalent. Why?

(c) Confirm part (b) by explicitly solving each of these systems.

5. The rank of the following matrices can be determined by inspection. Inspect these matrices and give their rank. Give reasons for your answers.

$$\text{(a)} \begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{(d)} \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$$

6. Find upper and lower bounds on the rank of the 4×3 matrix A, given that some system with coefficient matrix A has infinitely many solutions.

7. Answer True/False and explain your answers:

(a) Any homogeneous linear system with more unknowns than equations has a nontrivial solution.

(b) Any homogeneous linear system is consistent.

(c) If a linear system is inconsistent, then the rank of the augmented matrix exceeds the number of unknowns.

(d) Every matrix can be reduced to only one matrix in reduced row form.

(e) A system of 3 linear equations in 4 unknowns must have infinitely many solutions.

8. Suppose that $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and further that $a \neq 0$ and $ad - bc = 0$. Find the reduced row echelon form of A.

9. Give a rank condition for a homogeneous system that is equivalent to the system having a unique solution. Justify your answer.

10. Prove or disprove: if two linear systems are equivalent, then they must have the same size augmented matrix.

11. Use Theorem 1.4.8 to show that any two reduced row forms for a matrix A must have the same number of nonzero rows.

12. Suppose that the matrix C can be written in the augmented form $C = [A \,|\, B]$, where the matrix B may have more than one column. Prove that $\operatorname{rank} C \leq \operatorname{rank} A + \operatorname{rank} B$.

1.5. *Computational Notes and Projects

Roundoff Errors

In many practical problems, calculations are not exact. There are several reasons for this unfortunate fact. For one thing, scientific calculators are by their very nature only finite precision machines. That is, only a fixed number of significant digits of the numbers we are calculating may be used in any given calculation. For instance, verify this simple arithmetic fact on a calculator or computational software like MATLAB (but excluding computer algebra systems such as Derive, Maple or Mathematica – since symbolic calculation is the default on these systems, they will give the correct answer):

$$((\frac{2}{3} + 100) - 100) - \frac{2}{3} = 0$$

In many cases this calculation will not yield 0. The problem is that if, for example, a calculator uses 6 digit accuracy, then $2/3$ is calculated as 0.666667, which is really incorrect. Even if arithmetic calculations were exact, the data which form the basis of our calculations are often derived from scientific measurement which themselves will almost certainly be in error. Starting with erroneous data and doing an exact calculation can be as bad as starting with exact data and doing an inexact calculation. In fact, in a certain sense they are equivalent to each other. Error resulting from truncating data for storage or finite precision arithmetic calculations is called *roundoff error*.

We will not give an elaborate treatment of roundoff error. A thorough analysis can be found in the Golub and Van Loan text [**5**] of the bibliography, a text which is considered a standard reference work. This subject is a part of an entire field of applied mathematics known as numerical analysis. We will consider this question: could roundoff error be a significant problem in Gaussian elimination? It isn't at all clear that there is a problem. After all, even in the above example, the final error is relatively small. Is it possible that with all the arithmetic performed in Gaussian elimination the errors pile up and become large? The answer is "yes." With the advent of computers came a heightened interest in these questions. In the early 1950's numerical analysts intensified efforts to determine whether or not Gaussian elimination can reliably solve larger linear systems. In fact, we don't really have to look at complicated examples to realize that there is a potential difficulty. Consider the following example.

EXAMPLE 1.5.1. Let ϵ be a number so small that our calculator yields $1+\epsilon=1$. This equation appears a bit odd, but from the calculator's point of view it may be perfectly correct; if, for example, our calculator performs 6 digit arithmetic, then $\epsilon = 10^{-6}$will do nicely. Notice that with such a calculator, $1+1/\epsilon = (\epsilon+1)/\epsilon = 1/\epsilon$. Now solve the linear system

$$\begin{aligned} \epsilon x_1 + x_2 &= 1 \\ x_1 - x_2 &= 0 \end{aligned}$$

SOLUTION. Let's solve this system by Gauss-Jordan elimination with our calculator to obtain

$$\begin{bmatrix} \epsilon & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \xrightarrow[E_{21}(-1/\epsilon)]{} \begin{bmatrix} \epsilon & 1 & 1 \\ 0 & 1/\epsilon & -1/\epsilon \end{bmatrix} \xrightarrow[E_2(\epsilon)]{} \begin{bmatrix} \epsilon & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$\xrightarrow[E_{12}(-1)]{} \begin{bmatrix} \epsilon & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow[E_1(1/\epsilon)]{} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Thus we obtain the calculated solution $x_1 = 0$, $x_2 = 1$. This answer is spectacularly bad! If $\epsilon = 10^{-6}$ as above, then the correct answer is

$$x_1 = x_2 = \frac{1}{1+\epsilon} = 0.99999909999990\cdots$$

Our calculated answer is not even good to one digit. So we see that there can be serious problems with Gaussian or Gauss-Jordan elimination on finite precision machines. □

It turns out that information that would be significant for x_1 in first equation is lost in the truncated arithmetic that says that $1+1/\epsilon = 1/\epsilon$. There is a fix for problems such as this, namely a technique called *partial pivoting.* The idea is fairly simple: do not choose the next available column entry for a pivot. Rather, search down the column in Pivoting Strategies question for the largest entry (in absolute value). Then switch rows, if necessary, and use this entry as a pivot. For instance, in the preceding example, we would not pivot off the ϵ entry of the first column. Since the entry of the second row, first column, is larger in absolute value, we would switch rows and then do the usual Gaussian elimination step. Here is what we would get (remember that with our calculator $1 + \epsilon = 1$):

$$\begin{bmatrix} \epsilon & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix} \overrightarrow{E_{21}} \begin{bmatrix} 1 & -1 & 0 \\ \epsilon & 1 & 1 \end{bmatrix} \overrightarrow{E_{21}(-\epsilon)} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \overrightarrow{E_{12}(-\epsilon)} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Now we get the quite acceptable answer $x_1 = x_2 = 1$.

But partial pivoting is not a panacea for numerical problems. In fact, it can be easily defeated. Multiply the second equation by ϵ^2 and we get a system for which partial pivoting still picks the wrong pivot. Here the problem is a matter of scale. It can be cured by dividing each row by the largest entry of the row before beginning the Gaussian elimination process. This procedure is known as *row scaling*. The combination of row scaling and partial pivoting overcomes many of the numerical problems of Gaussian or Gauss-Jordan elimination (but not all!). There is a more drastic procedure, known as *complete pivoting*. In this procedure one searches all the unused rows (excluding the right hand sides) for the largest entry, then uses it as a pivot for Gaussian elimination. The columns used in this procedure do not move in that left-to-right fashion we are used to seeing in system solving. It can be shown rigorously that the error of roundoff propagates in a predictable and controlled fashion with complete pivoting; in contrast, we do not really have a satisfactory explanation as to why row scaling and partial pivoting work well. Yet in most cases they do reasonably well. Since this combination involves much less calculation than complete pivoting, it is the method of choice for many problems.

There are deeper reasons for numerical problems in solving some systems than the one the preceding example illustrates. One difficulty has to do with the "sensitivity" of the coefficient matrix to small changes. That is, in some systems, small changes in the coefficient matrix lead to dramatic changes in the *exact* answer. The practical effect of roundoff error can be shown to be equivalent to introducing small changes in the coefficient matrix and obtaining an exact answer to the perturbed (changed) system. There is no cure for these difficulties. A classical example of this type of problem, the Hilbert matrix, is discussed in one of the projects below. We will attempt to quantify this "sensitivity" in Chapter 6.

Computational Efficiency of Gaussian Elimination

How much work is it to solve a linear system and how does the amount of work grow with the dimensions of the system? The first thing we need is a unit of work. In computer science one of the principal units of work is a *flop* (floating point operation), namely a single $+, -, \times$, or $\div$. For example, we say the amount of work in computing $e + \pi$ or $e \times \pi$ is one flop, while the work in calculating $e + 3 \times \pi$ is two flops. The following example is extremely useful.

EXAMPLE 1.5.2. How many flops does it cost to add a multiple of one row to another, as in Gaussian elimination, given that the rows have n elements each?

SOLUTION. A little experimentation with an example or two shows that that the answer should be $2n$. Here is a justification of that count. Say that row $\mathbf{a}$ is to be multiplied by the scalar α, and added to the row $\mathbf{b}$. Designate the row $a = [a_i]$ and the row $b = [b_i]$. We have n entries to worry about. Consider a typical one, say the ith one. The ith entry of $\mathbf{b}$, namely b_i, will be replaced by the quantity

$$b_i + \alpha a_{i.}$$

The amount of work in this calculation is two flops. Since there are n entries to compute, the total work is $2n$ flops. □

Our goal is to determine the expense of solving a system by Gauss-Jordan elimination. For the sake of simplicity, let's assume that the system under consideration has n equations in n unknowns and the coefficient matrix has rank n. This ensures that we will have a pivot in every row of the matrix. We won't count row exchanges either, since they don't involve any flops. (We should remark that this may not be realistic on a fast computer, since memory fetches and stores may not take significantly less time than a floating point operation.) Now consider the expense of clearing out the entries under the first pivot. A picture of the augmented matrix looks something like this, where an '$\times$' is an entry which may not be 0 and an '$\times$' is a nonzero pivot entry:

$$\begin{bmatrix} \times & \times & \cdots & \times \\ \times & \times & \cdots & \times \\ \vdots & \vdots & \vdots & \vdots \\ \times & \times & \cdots & \times \end{bmatrix} \xrightarrow[\text{el.ops}]{n-1} \begin{bmatrix} \times & \times & \cdots & \times \\ 0 & \times & \cdots & \times \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \times & \cdots & \times \end{bmatrix}$$

Each elementary operation will involve adding a multiple of the first row, starting with the *second* entry, since we don't need to do arithmetic in the first column – we know what goes there, to the $n-1$ subsequent rows. By the preceding example, each of these elementary operations will cost $2n$ flops. Add 1 flop for the cost of determining the multiplier to obtain $2n+1$. So the total cost of zeroing out the first column is $(n-1)(2n+1)$ flops. Now examine the lower unfinished block in the above figure. Notice that it's as though we were starting over with the row and column dimensions reduced by 1. Therefore, the total cost of the next phase is $(n-2)(2(n-1)+1)$ flops. Continue in this fashion and we obtain a count of

$$0 + \sum_{j=2}^{n}(j-1)(2j+1) = \sum_{j=1}^{n}(j-1)(2j+1) = \sum_{j=1}^{n} 2j^2 - j - 1$$

flops. Recall the identities for sums of consecutive integers and their squares:

$$\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$$

$$\sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}$$

Thus we have a total flop count of

$$\sum_{j=1}^{n} 2j^2 - 3j + 1 = 2\frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)}{2} - n = \frac{2n^3}{3} + \frac{n^2}{2} - \frac{7n}{6}$$

This is the cost of forward solving. Now let's simplify our answer a bit more. For large n we have that n^3 is much larger than n or n^2 (e.g., for $n = 10$ compare 1000 to 10 or 100). Hence, we ignore the lower degree terms and arrive at a simple approximation to the number of flops required to forward solve a linear system of n equations in n unknowns using Gauss-Jordan elimination. There remains the matter of back solving. We leave as an exercise to show that the total work of back solving is quadratic in n. Therefore the "leading order" approximation which we found for forward solving remains unchanged.

THEOREM 1.5.3. *The number of flops required to solve a linear system of n equations in n unknowns using Gaussian or Gauss-Jordan elimination without row exchanges is* approximately $2n^3/3$.

Thus, for example, the work of forward solving a system of 21 equations in 21 unknowns is approximately $2 \cdot 21^3/3 = 6174$ flops. Compare this to the exact answer of 6374.

Project Topics

In this section we give a few samples of project material. These projects provide an opportunity for students to explore a subject in a greater depth than exercises permit. They also provide an opportunity for students to hone their scientific computing and writing skills. They are well suited to team effort, and writing expectations can range from a summary of the answers to a detailed report on the project. The first sample project is written with the expectation of a fairly elaborate project report. The instructor has to define her/his own expectations for students. Likewise, the computing platform used for the projects will vary. We cannot discuss every platform in this text, so we will give examples of implementation notes that an instructor might supply for a few of them. The instructor will have to modify that portion of the project to match the local configuration and provide additional background about the use of the computing platform.

Notes to students about project/reports: The first thing you need to know about report writing is the intended audience. Usually, you may assume that your report will be read by your supervisors, who are technical people such as yourself. Therefore, you should write a brief statement of the problem and discussion of methodology. In practice reports, you assume physical laws and assumptions without further justification, but in real life you would be expected to offer some explanation of physical principles you employ in constructing your model. Another good point to have in mind is a target length for your paper. Do not clutter your work with long lists of numbers and try to keep the length at a minimum rather than maximum. Most kinds of discourse should have three parts: a beginning, a middle and an end. Roughly, a beginning should consist of introductory material. In the middle you develop the ideas described or theses proposed in the introduction, and in the end you summarize your work and tie up loose ends.

Of course, rules about paper writing are not set in concrete, and authors vary on exactly how they organize papers of a given kind. Also, a part could be quite short; for example, an introduction might only be a paragraph or two. Here is a sample skeleton for a report (perhaps rather more elaborate than you need): 1. Introduction (title page, summary and conclusions), 2. Main Sections (problem statement, assumptions, methodology, results, conclusions), 3. Appendices (such as mathematical analysis, graphs, possible extensions, etc.) and References.

A few additional notes: Pay attention to appearance and neatness, but don't be overly concerned about your writing style. A good rule to remember is "Simpler is better." Prefer short and straightforward sentences to convoluted ones. Use a vocabulary that you are comfortable with. Be sure to use a spell-checker if one is available to you.

A given project/report assignment may be supplied with a report template by your instructor or carry explicit instructions about format, intended audience, etc. It is important to read and follow these instructions carefully. Naturally, such instructions would take precedence over any of the preceding remarks.

The first two of these projects are based on the material of Section 1.1 in relation to diffusion processes.

Project: Heat Flow I

Description of the problem: You are working for the firm Universal Dynamics on a project which has a number of components. You have been assigned the analysis of a component which is similar to a laterally insulated rod. The problem you are concerned with is as follows: part of the specs for the rod dictate that *no point of the rod should stay at temperatures above* 60 *degrees Celsius for a long period of time.* You must decide if any of the materials listed below are acceptable for making the rod and write a report on your findings. You may assume that the rod is of unit length. Suppose further that internal heat sources come from a position dependent function $f(x)$, $0 \leq x \leq 1$ and that heat is also generated at each point in amounts proportional to the temperature at the point. Also suppose that the left and right ends of the rod are held at 0 and 50 degrees Celsius, respectively. When sufficient time passes, the temperature of the rod at each point will settle down to "steady state" values, dependent only on position x. These are the temperatures you are interested in. Refer to the discussion in 1.1 for the details of the descriptive equations that result from discretizing the problem into finitely many nodes. Here k is the *thermal conductivity* of the rod, which is a property associated with the material used to make the rod. For your problem take the source term to be $f(x) = 200\cos(x^2)$. Here are the conductivity constants for the materials with which your company is considering building the rod. Which of these materials (if any) are acceptable?

Platinum: $k = .17$

Zinc: $k = .30$

Aluminum: $k = .50$

Gold: $k = .75$

Silver: $k = 1.00$

Procedure: For the solution of the problem, formulate a discrete approximation to the BVP just as in Example 1.1.5. Choose an integer n and divide the interval [0, 1] into $n+1$ equal subintervals with endpoints $0 = x_0, x_1, ..., x_{n+1} = 1$. Then the width of each subinterval is $h = 1/(n+1)$. Next let u_i be our approximation to $u(x_i)$ and proceed as in Example 1.1.5 . There results a linear system of n equations in the n unknowns $u_1, u_2, ..., u_n$. For this problem divide the rod into 4 equally sized subintervals and take $n = 3$. Use the largest u_i as an estimate of the highest temperature at any point in the rod. Now double the number of subintervals and see if your values for u change appreciably at a given value of x. If they do, you may want to repeat this procedure until you obtain numbers that you judge to be satisfactory.

Implementation Notes (for users of Mathematica): Set up the coefficient matrix a and right hand side b for the system. Both the coefficient matrix and the right hand side can be set up using the Table command of Mathematica. For b, the command `100* h^2*Table[Cos[(i h)^ 2,{i,n}]/k` will generate b, except for the last coordinate. Use the command `b[[14]] = b[[14]] + 50` to add $u(1)$ to the right hand side of the system and get the correct b. For a: the command `Table[Switch[i-j,1,-1,0,2,-1,-1,_,0],{i,n},{j,n}]` will generate a matrix of the desired form. (Use the Mathematica on line help for all commands you want to know more about.) For floating point numbers: we want to simulate ordinary floating point calculations on Mathematica. You will get some symbolic expressions which we don't want, e.g., for b. To turn b into floating point approximation, use the command `b = N[b]`. The N[ ] function turns the symbolic values of b into numbers, with a precision of about 16 digits if no precision is specified. For solving linear systems use the command `u = LinearSolve[a,b]`, which will solve the system with coefficient matrix a and right hand side b, and store the result in u. About vectors: Mathematica does not distinguish between row vectors or column vectors unless you insist on it. Hardcopy: You can get hardcopy from Mathematica. Be sure to make a presentable solution for the project. You should describe the form of the system you solved and at least summarize your results. This shouldn't be a tome (don't simply print out a transcript of your session), nor should it be a list of numbers.

Project: Heat Flow II

Problem Description: You are given a laterally insulated rod of a homogeneous material whose conductivity properties are unknown. The rod is laid out on the x-axis, 0 <= x <= 1. A current is run through the rod, which results in a heat source of 10 units of heat (per unit length) at each point along the rod. The rod is held at zero temperature at each end. After a time the temperatures in the rod settle down to a steady state. A single measurement is taken at x=0.3 which results in a temperature reading of approximately 11 units. Based on this information, determine the best estimate you can for the true value of the conductivity constant k of the material. Also try to guess a formula for the shape of the temperature function on the interval [0,1] that results when this value of the conductivity is used.

Methodology: You should use the model that is presented on pages 4-6 of the text. This will result in a linear system, which Maple can solve. One way to proceed is simply to use trial and error until you think you've hit on the right value of k, that is, the one that gives a value of approximately 11 units at x =0.3. Then plot the resulting approximate function doing a dot-to-dot on the node values. You should give some thought to step size h.

Output: Return your results in the form of an annotated Maple notebook, which should have the name of the team members at the top of the file and an explanation of your solution in text cells interspersed between input cells, that the user can happily click his/her way through. This explanation should be intelligible to your fellow students.

Comments: This project introduces you to a very interesting area of mathematics called "inverse theory." The idea is, rather than proceeding from problem (the governing equations for temperature values) to solution (temperature values), you are given the "solution", namely the measured solution value at a point, and are to determine from this information the "problem", that is, the conductivity coefficient that is needed to define the governing equations.

Project: The Accuracy of Gaussian Elimination

Description of the problem: This project is concerned with determining the accuracy of Gaussian elimination as applied to two linear systems, one of which is known to be difficult to solve numerically. Both of these systems will be square (equal number of unknowns and equations) and have a unique solution. Also, both of these systems are to be solved for various sizes, namely $n = 4, 8, 12, 16$. In order to get a handle on the error, our main interest, we shall start with a known answer. The answer shall consist of setting all variables equal to 1. So it is the solution vector $(1, 1, \ldots, 1)$. The coefficient matrix shall be one of two types:

(1) A Hilbert matrix, i.e., an $n \times n$ matrix given by the formula

$$H_n = \left[\frac{1}{i+j-1}\right]$$

(2) An $n \times n$ matrix with random entries between 0 and 1.

The right hand side vector b is uniquely determined by the coefficient matrix and solution. In fact, the entries of b are easy to obtain: simply add up all the entries in the ith row of the coefficient matrix to obtain the ith entry of b.

The problem is to measure the error of Gaussian elimination. This is done by finding the largest (in absolute value) difference between the computed value of each variable and actual value, which in all cases is 1. Discuss your results and draw conclusions from your experiments.

Implementation Notes (for users of Maple): Maple has a built-in procedure for defining a Hilbert matrix A of size n, as in the command `A := hilbert(n);`. Before executing this command (and most other linear algebra commands), you must load the linear algebra package by the command `with(linalg);`. A vector of 1's of size n can also be constructed by the single command `x := vector(n,1);`. To multiply this matrix and vector together use the command `evalm(A &* x);`. There is a feature that all computer algebra systems have: they do exact arithmetic whenever possible. Since we are trying to gauge the effects of finite precision calculations, we don't want exact answers (such as 425688/532110), but rather, finite precision floating point answers (such as 0.8). Therefore, it would be a good idea at some point to force the quantities in question to be finite precision numbers by encapsulating their definitions in an evaluate as floating point command, e.g., `evalf(evalm(A &* x));`. This will force the CAS to do finite precision arithmetic.

1.5 Exercises

1. Carry out the calculation $((\frac{2}{3}+100)-100)-\frac{2}{3}$ on a scientific calculator. Do you get the correct answer?

2. Enter the matrix A given below into a computer algebra system and use the available commands to compute (a) the rank of A and (b) the reduced row echelon form of A. (For example, in Maple the relevant commands are `rref(A)` and `rank(A)`.) Now convert A into its floating point form and execute the same commands. Do you get the same answers? If not, which is correct?

$$A=\begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 6 & 18 & -15 & -6 & 12 & -9 & -3 \\ 0 & 0 & 5 & 10 & 0 & 15 & 5 \\ 2 & 6 & 0 & 8 & 4 & 18 & 6 \end{bmatrix}$$

3. Show that the flop count for back solving an $n \times n$ system is quadratic in n. *Hint:* At the jth stage the total work is $j+2\left[(n-1)+(n-2)+\ldots+(n-j)\right]$.

Review

Chapter 1 Exercises

1. Calculate the following:

(a) $|2+4i|$ (b) $-7i^2+6i^3$ (c) $(3+4i)(7-6i)$ (d) $\overline{7-6i}$

2. Solve the following systems for the (complex) variable z. Express your answers in standard form where possible.

(a) $(2+i)z=4-2i$ (b) $z^4=-16$ (c) $z+1/z=1$ (d) $(z+1)(z^2+1)=0$

3. Find the polar and standard form of the complex numbers

(a) $1/(1-i)$ (b) $-2e^{i\pi/3}$ (c) $i(i+\sqrt{3})$ (d) $-1+i$ (e) $ie^{\pi/4}$

4. The following are augmented matrices of linear systems. In each case, reduce the matrix to reduced row echelon form and exhibit the solution(s) to the system.

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

5. Consider the linear system:

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

(a) Solve this system by reducing the augmented matrix to reduced row echelon form. Show all row operations.

(b) Find the rank of the augmented matrix and express its reduced row echelon form as a product of elementary matrices and the augmented matrix.

(c) This system will have solutions for any right hand side. Justify this fact in terms of rank.

6. Fill in the blanks or answer True/False with justification for your answer:

(a) If A is a 3×7 matrix then the rank of A is at most ____________

(b) If A is a 4×8 matrix, then the nullity of A could be larger than 4 (T/F):

(c) Any homogeneous (right hand side vector 0) linear system is consistent (T/F):

(d) The rank of a nonzero 3×3 matrix with all entries equal is ________

(e) Some polynomial equations $p(z) = 0$ have no solutions z (T/F):

7. What is the locus in the plane of complex numbers z such that $|z + 3| = |z - 1|$?

8. For what values of c are the following systems inconsistent, with unique solution or with infinitely many solutions?

(a) $$\begin{aligned} cx_1 + x_2 + x_3 &= 2 \\ x_1 + cx_2 + x_3 &= 2 \\ x_1 + x_2 + cx_3 &= 2 \end{aligned}$$ (b) $$\begin{aligned} x_1 + 2x_2 - x_1 &= c \\ x_1 + 3x_2 + x_3 &= 1 \\ 3x_1 + 7x_2 - x_3 &= 4 \end{aligned}$$ (c) $$\begin{aligned} x_2 + cx_3 &= 0 \\ x_1 - cx_2 &= 1 \end{aligned}$$

9. Show that a system of linear equations has a unique solution if and only if every column, except the last one, of the reduced row echelon form of the augmented matrix has a pivot entry in it.

CHAPTER 2

MATRIX ALGEBRA

In Chapter 1 we used matrices and vectors as simple storage devices. In this chapter matrices and vectors take on a life of their own. We develop the arithmetic of matrices and vectors. Much of what we do is motivated by a desire to extend the ideas of ordinary arithmetic to matrices. Our notational style of writing a matrix in the form $A = [a_{ij}]$ hints that a matrix could be treated like a single number. What if we could manipulate equations with matrix and vector quantities in the same way that we do scalar equations? We shall see that this is a useful idea. Matrix arithmetic gives us new powers for formulating and solving practical problems. In this chapter we will use it to find effective methods for solving linear and nonlinear systems, solve problems of graph theory and analyze an important modeling tool of applied mathematics called a Markov chain.

2.1. Matrix Addition and Scalar Multiplication

To begin our discussion of arithmetic we consider the matter of equality of matrices. Suppose that A and B represent two matrices. When do we declare them to be equal? The answer is, of course, if they represent the same matrix! Thus we expect that all the usual laws of equalities will hold (e.g., equals may be substituted for equals) and in fact, they do. There are times, however, when we need to prove that two symbolic matrices must be equal. For this purpose, we need something a little more precise. So we have the following definition, which includes vectors as a special case of matrices.

DEFINITION 2.1.1. Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are said to be *equal* if these matrices have the same size, and for each index pair (i, j), $a_{ij} = b_{ij}$, that is, corresponding entries of A and B are equal.

EXAMPLE 2.1.2. Which of the following matrices are equal, if any?

$$\text{(a)} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 0 & 0 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix} \quad \text{(d)} \begin{bmatrix} 0 & 1 \\ 1-1 & 1+1 \end{bmatrix}$$

SOLUTION. The answer is that only (c) and (d) have any chance of being equal, since they are the only matrices in the list with the same size (2×2). As a matter of fact, an entry by entry check verifies that they really are equal. $\square$

Matrix Addition and Subtraction

How should we define addition or subtraction of matrices? We take a clue from elementary two and three dimensional vectors, such as the type we would encounter in geometry or calculus. There, in order to add two vectors, one condition had to hold: the vectors had to be the same size. If they were the same size, we simply added the vectors coordinate by coordinate to obtain a new vector of the same size. That is precisely what the following definition says.

DEFINITION 2.1.3. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $m \times n$ matrices. Then the *sum* of the matrices, denoted as $A + B$, is the $m \times n$ matrix defined by the formula

$$A + B = [a_{ij} + b_{ij}]$$

The *negative* of the matrix A, denoted as $-A$, is defined by the formula

$$-A = [-a_{ij}]$$

Finally, the *difference* of A and B, denoted as $A - B$, is defined by the formula

$$A - B = [a_{ij} - b_{ij}]$$

Notice that matrices must be the same size before we attempt to add them. We say that two such matrices or vectors are *conformable for addition*.

EXAMPLE 2.1.4. Let

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

Find $A + B$, $A - B$, and $-A$.

SOLUTION. Here we see that

$$\begin{aligned} A + B &= \begin{bmatrix} 3 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} + \begin{bmatrix} -3 & 2 & 1 \\ 1 & 4 & 0 \end{bmatrix} = \begin{bmatrix} 3-3 & 1+2 & 0+1 \\ -2+1 & 0+4 & 1+0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 3 & 1 \\ -1 & 4 & 1 \end{bmatrix} \end{aligned}$$

Likewise,

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

The negative of A is even simpler:

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

□

Scalar Multiplication

The next arithmetic concept we want to explore is that of scalar multiplication. Once again, we take a clue from the elementary vectors, where the idea behind scalar multiplication is simply to "scale" a vector a certain amount by multiplying each of its coordinates by that amount. That is what the following definition says.

DEFINITION 2.1.5. Let $A = [a_{ij}]$ be an $m \times n$ matrix and c a scalar. Then the *scalar product* of the scalar c with the matrix A, denoted by cA, is defined by the formula

$$cA = [ca_{ij}].$$

Recall that the default scalars are real numbers, but they could also be complex numbers.

EXAMPLE 2.1.6. Let

$$A = \begin{bmatrix} 3 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}, \quad \text{and} \quad c = 3$$

Find cA, $0A$, and $-1A$.

SOLUTION. Here we see that

$$cA = 3 \begin{bmatrix} 3 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 3 & 3 \cdot 1 & 3 \cdot 0 \\ 3 \cdot -2 & 3 \cdot 0 & 3 \cdot 1 \end{bmatrix} = \begin{bmatrix} 9 & 3 & 0 \\ -6 & 0 & 3 \end{bmatrix}$$

while

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

and

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

□

Linear Combinations

Now that we have a notion of scalar multiplication and addition, we can blend these two ideas to yield a very fundamental notion in linear algebra, that of a *linear combination*.

Linear Combinations

DEFINITION 2.1.7. A *linear combination*of the matrices $A_1, A_2, \ldots, A_n$ is an expression of the form

$$c_1 A_1 + c_2 A_2 + \ldots + c_n A_n$$

where $c_1, c_2, \ldots, c_n$ are scalars and $A_1, A_2, \ldots, A_n$ are matrices all of the same size.

EXAMPLE 2.1.8. Given that

$$A_1 = \begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}, \quad \text{and} \quad A_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},$$

compute the linear combination $-2A_1 + 3A_2 - 2A_3$.

SOLUTION. The solution is that

$$-2A_1+3A_2-2A_3=-2\begin{bmatrix}2\\6\\4\end{bmatrix}+3\begin{bmatrix}2\\4\\2\end{bmatrix}-2\begin{bmatrix}1\\0\\-1\end{bmatrix}$$
$$=\begin{bmatrix}-2\cdot 2+3\cdot 2-2\cdot 1\\-2\cdot 6+3\cdot 4-2\cdot 0\\-2\cdot 4+3\cdot 2-2\cdot(-1)\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

□

It seems like too much work to write out objects such as the vector $\begin{bmatrix}0\\0\\0\end{bmatrix}$ that occurred in the last equation; after all, we know that all the entries are all 0. So we make the following convention for convenience.

NOTATION 2.1.9. A *zero matrix* is a matrix whose every entry is 0. We shall denote such matrices by the symbol 0.

We have to be a bit careful, since this convention makes the symbol 0 ambiguous, but the meaning of the symbol will be clear from context, and the convenience gained is worth the potential ambiguity. For example, the equation of the preceding example is stated very simply as $-2A_1+3A_2-2A_3=0$, where we understand from context that 0 has to mean the 3×1 column vector of zeros. If we use boldface for vectors, we will also then use boldface for the vector zero, so some distinction is regained.

EXAMPLE 2.1.10. Suppose that a linear combination of matrices satisfies the identity $-2A_1+3A_2-2A_3=0$, as in the preceding example. Use this fact to express A_1 in terms of A_2 and A_3.

SOLUTION. To solve this example, just forget that the quantities A_1, A_2, A_3 are anything special and use ordinary algebra. First, add $-3A_2+2A_3$ to both sides to obtain

$$-2A_1+3A_2-2A_3-3A_2+2A_3=-3A_2+2A_3$$

so that

$$-2A_1=-3A_2+2A_3$$

and then multiplying both sides by the scalar $-1/2$ yields the identity

$$A_1=\frac{-1}{2}(-2A_1)=\frac{-3}{2}A_2+A_3$$

□

The linear combination idea has a really interesting application to linear systems, namely, it gives us another way to express the solution set of a linear system that clearly identifies the role of free variables. The following example illustrates this point.

EXAMPLE 2.1.11. Suppose that a linear system in the unknowns x_1, x_2, x_3, x_4 has general solution $(x_2+3x_4, x_2, 2x_1-x_4, x_4)$, where the variables x_2, x_4 are free. Describe the solution set of this linear system in terms of linear combinations with free variables as coefficients.

SOLUTION.The trick here is to use only the parts of the general solution involving x_2 for one vector and the parts involving x_4 as the other vectors in such a way that these vectors add up to the general solution. In our case we have

$$\begin{bmatrix} x_2 + 3x_4 \\ x_2 \\ 2x_1 - x_4 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_2 \\ 2x_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 3x_4 & \\ 0 & [r] \\ -x_4 & \\ x_4 & \end{bmatrix} = x_2 \begin{bmatrix} 1 \\ 1 \\ 2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 3 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

Now simply define vectors $A_1 = (1, 1, 2, 0)$, $A_2 = (3, 0, -1, 1)$ and we see that, since x_2 and x_4 are arbitrary, the solution set is

$$S = \{x_2 A_1 + x_4 A_2 \mid x_2, x_4 \in \mathbb{R}\}$$

In other words, the solution set to the system is the set of all possible linear combinations of the vectors A_1 and A_2. □

The idea of solution sets as linear combinations is an important one that we will return to in later chapters. You might notice that once we have the general form of a solution vector it's easy to determine the constant vectors A_1 and A_2. Simply set $x_2 = 1$ and the other free variable(s) – in this case just x_4 – to get the solution vector A_1, and set $x_4 = 1$ and $x_2 = 0$ to get the solution vector A_2.

Laws of Arithmetic

The last example brings up an important point: to what extent can we rely on the ordinary laws of arithmetic and algebra in our calculations with matrices and vectors? We shall see later in this section that for matrix *multiplication* there are some surprises. On the other hand, the laws for addition and scalar multiplication are pretty much what we would expect them to be. Taken as a whole, these laws are very useful; so much so that later in this text they will be elevated to the rank of an axiom system for what are termed "vector spaces." Here are the laws with their customary names. These same names can apply to more than one operation. For instance, there is a closure law for addition and one for scalar multiplication as well.

Laws of Matrix Addition and Scalar Multiplication. Let A, B, C be matrices of the same size $m \times n$, 0 the $m \times n$ zero matrix, and c and d scalars. Then

1. (Closure Law) $A + B$ is an $m \times n$ matrix.
2. (Associative Law) $(A + B) + C = A + (B + C)$
3. (Commutative Law) $A + B = B + A$
4. (Identity Law) $A + 0 = A$
5. (Inverse Law) $A + (-A) = 0$
6. (Closure Law) cA is an $m \times n$ matrix.
7. (Associative Law) $c(dA) = (cd)A$
8. (Distributive Law) $(c + d)A = cA + dA$
9. (Distributive Law) $c(A + B) = cA + cB$
10. (Monoidal Law) $1A = A$

It is fairly straightforward to prove from definitions that these laws are valid. The verifications all follow a similar pattern, which we illustrate by verifying the commutative law for addition: let $A = [a_{ij}]$ and $B = [b_{ij}]$ be given $m \times n$ matrices. Then we have that

$$\begin{aligned} A + B &= [a_{ij} + b_{ij}] \\ &= [b_{ij} + a_{ij}] \\ &= B + A \end{aligned}$$

where the first and third equalities come from the definition of matrix addition, and the second equality follows from the fact that for all indices i and j, $a_{ij} + b_{ij} = b_{ij} + a_{ij}$ by the commutative law for addition of scalars.

2.1 Exercises

1. Calculate

(a) $\begin{bmatrix} 1 & 2 & -1 \\ 0 & 2 & 2 \end{bmatrix} - \begin{bmatrix} 3 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ (b) $2\begin{bmatrix} 1 \\ 3 \end{bmatrix} - 5\begin{bmatrix} 2 \\ 2 \end{bmatrix} + 3\begin{bmatrix} 4 \\ 1 \end{bmatrix}$

(c) $2\begin{bmatrix} 1 & 4 \\ 0 & 0 \end{bmatrix} + 3\begin{bmatrix} 0 & 0 \\ 2 & 1 \end{bmatrix}$ (d) $a\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} + b\begin{bmatrix} -1 & 0 \\ -1 & 1 \end{bmatrix}$

(e) $2\begin{bmatrix} 1 & 2 & -1 \\ 0 & 0 & 2 \\ 0 & 2 & 2 \end{bmatrix} - 3\begin{bmatrix} 3 & 1 & 0 \\ 5 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ (f) $x\begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} - 5\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} + y\begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix}$

2. Let $A = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 2 \\ 1 & -2 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 2 \end{bmatrix}$, and compute the following, where possible.

(a) $A + 3B$ (b) $2A - 3C$ (c) $A - C$ (d) $6B + C$ (e) $2C - 3(A - 2C)$

3. With A, C as in Exercise 2, solve the equations (a) $2X + 3A = C$ and (b) $A - 3X = 3C$ for the unknown X.

4. Show how to write each of the following vectors as a linear combination of constant vectors with scalar coefficients x, y or z.

(a) $\begin{bmatrix} x + 2y \\ 2x - z \end{bmatrix}$ (b) $\begin{bmatrix} x - y \\ 2x + 3y \end{bmatrix}$ (c) $\begin{bmatrix} 3x + 2y \\ -z \\ x + y + 5z \end{bmatrix}$ (d) $\begin{bmatrix} x - 3y \\ 4x + z \\ 2y - z \end{bmatrix}$

5. Find scalars a, b, c such that

$$\begin{bmatrix} c & b \\ 0 & c \end{bmatrix} = \begin{bmatrix} a - b & c + 2 \\ a + b & a - b \end{bmatrix}$$

6. Show that any matrix of the form $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ can be expressed as a linear combination of the four matrices $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, and $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.

7. Verify that the associative law and both distributive laws for addition hold for $c = 2$, $d = -3$ and

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

8. Prove that the associative law for addition of matrices (page 53) holds.

9. Prove that both distributive laws (page 53) hold.

10. Prove the following assertions for $m \times n$ matrices A and B by using the laws of matrix addition and scalar multiplication. Clearly specify each law that you use.

(a) If $cA = 0$ for some scalar c, then either $c = 0$ or $A = 0$.

(b) If $B = cB$ for some scalar $c \neq -1$, then $B = 0$.

2.2. Matrix Multiplication

Matrix multiplication is somewhat more subtle than matrix addition and scalar multiplication. Of course, we could define matrix multiplication to be coordinate-wise, just as addition is. But our motivation is not merely to make definitions, but rather to make *useful* definitions.

Definition of Multiplication

To motivate the definition, let us consider a single linear equation

$$2x - 3y + 4z = 5.$$

We will find it handy to think of the left hand side of the equation as a "product" of the coefficient matrix $[2, -3, 4]$ and the column matrix of unknowns $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$. Thus, we have that the product of this row and column is

$$[2, -3, 4] \begin{bmatrix} x \\ y \\ z \end{bmatrix} = [2x - 3y + 4z].$$

Notice that we have made the result of the product into a matrix (in this case 1×1). This introduces us to a permanent abuse of notation that is almost always used in linear algebra: we don't distinguish between the scalar a and the 1×1 matrix $[a]$, though technically perhaps we should. In the same spirit, we make the following definition.

DEFINITION 2.2.1. The *product* of the $1 \times n$ row $[a_1, a_2, ..., a_n]$ with the $n \times 1$ column $\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ is defined to be the 1×1 matrix $[a_1 b_1 + a_2 b_2 + ... + a_n b_n]$.

It is this row-column product strategy that guides us to the general definition. Notice how the column number of the first matrix had to match the row number of the second, and that this number disappears in the size of the resulting product. This is exactly what happens in general.

DEFINITION 2.2.2. Let $A = [a_{ij}]$ be an $m \times p$ matrix and $B = [b_{ij}]$ be a $p \times n$ matrix. Then the *product* of the matrices A and B, denoted by $A \cdot B$ (or simply AB), is the $m \times n$ matrix whose (i, j)th entry, for $1 \leq i \leq m$ and $1 \leq j \leq n$, is the entry of the product of the ith row of A and the jth column of B; more specifically, the (i, j)th entry of AB is

$$a_{i1} b_{1j} + a_{i2} b_{2j} + ... + a_{ip} b_{pj}$$

Notice that, unlike addition, two matrices may be of different sizes when we can multiply them together. If A is $m \times p$ and B is $p \times n$, we say that A and B are *conformable* for multiplication. It is also worth noticing that if A and B are *square* and of the same size, then the products AB and BA are always defined.

Some Illustrative Examples

Let's check our understanding with a few examples.

EXAMPLE 2.2.3. Compute, if possible, the products of the following pairs of matrices A, B.

$$(a) \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix}, \quad \begin{bmatrix} 4 & -2 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \qquad (b) \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & -1 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$(c) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix} \qquad (d) \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 \end{bmatrix}$$

$$(e) \begin{bmatrix} 1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad (f) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$$

SOLUTION. In problem (a) A is 2×3 and B is 3×2. First check conformability for multiplication. Stack these dimensions along side each other and see that the 3's match; now "cancel" the matching middle 3's to obtain that the dimension of the product is $2 \times \cancel{3}\ \cancel{3} \times 2 = 2 \times 2$. To obtain, for example, the $(1, 2)$th entry of the product matrix multiply the first row of A and second column of B to obtain

$$[1, 2, 1] \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = [1 \cdot (-2) + 2 \cdot 1 + 1 \cdot 1] = [1]$$

The full product calculation looks like this:

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix} \begin{bmatrix} 4 & -2 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1\cdot 4 + 2\cdot 0 + 1\cdot 2 & 1\cdot(-2) + 2\cdot 1 + 1\cdot 1 \\ 2\cdot 4 + 3\cdot 0 + (-1)\cdot 2 & 2\cdot(-2) + 3\cdot 1 + (-1)\cdot 1 \end{bmatrix}$$
$$= \begin{bmatrix} 6 & 1 \\ 6 & -2 \end{bmatrix}$$

A size check of part (b) reveals a mismatch between the column number of the first matrix (3) and the row number (2) of the second matrix. Thus these matrices are *not conformable* for multiplication in the specified order. Hence

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

is undefined.

Things work better in (c), where the size check gives $2\times \cancel{2}\ \cancel{2}\times 3 = 2\times 3$ as the size of the product. As a matter of fact, this is a rather interesting calculation:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 0\cdot 2 & 1\cdot 2 + 0\cdot 3 & 1\cdot 1 + 0\cdot(-1) \\ 0\cdot 1 + 1\cdot 2 & 0\cdot 2 + 1\cdot 3 & 0\cdot 1 + 1\cdot(-1) \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix}$$

Notice that we end up with the second matrix in the product. This is similar to the arithmetic fact that $1\cdot x = x$ for a given real number x. So the matrix on the left acted like a multiplicative identity. We will see later that this is no accident.

In problem (d) a size check shows that the product has size $2\times \cancel{1}\ \cancel{1}\times 2 = 2\times 2$. The calculation gives

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 0\cdot 1 & 0\cdot 2 \\ 0\cdot 1 & 0\cdot 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

For part (e) the size check shows gives $1\times \cancel{2}\ \cancel{2}\times 1 = 1\times 1$. Hence the product exists and is 1×1. The calculation gives

$$\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = [1\cdot 0 + 2\cdot 0] = [0]$$

Matrix Multiplication Not Commutative or Cancellative

Something very interesting comes out of this calculation. Notice that for this choice of A and B we have that AB and BA are *not* the same matrices - never mind that their entries are all 0's - the important point is that these matrices are not even the same size! Thus a very familiar law of arithmetic, the commutativity of multiplication, has just fallen by the wayside.

Finally, for the calculation in (f), notice that

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

There's something very curious here, too. Notice that two nonzero matrices of the same size multiplied together to give a zero matrix. This kind of thing never happens in ordinary arithmetic, where the cancellation law assures that if $a\cdot b = 0$ then $a = 0$ or $b = 0$. $\square$

The calculation in (c) inspires some more notation. The left-hand matrix of this product has a very important property. It acts like a "1" for matrix multiplication. So it deserves its own name.

NOTATION 2.2.4. A matrix of the form

Identity Matrix

$$I_n = \begin{bmatrix} 1 & 0 & \dots & & 0 \\ 0 & 1 & 0 & & \\ \vdots & & \ddots & & \\ & & & 1 & 0 \\ 0 & & & 0 & 1 \end{bmatrix} = [\delta_{ij}]$$

is called an $n \times n$ *identity matrix*. The (i, j)th entry of I_n is designated by the Kronecker symbol δ_{ij} which is 1 if $i = j$ and 0 otherwise. If n is clear from context, we simply write I in place of I_n.

So we see in the previous example that the left hand matrix of part (c) is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

Linear Systems as a Matrix Product

Let's have another look at a system we examined in Chapter 1. We'll change the names of the variables from x, y, z to x_1, x_2, x_3 in anticipation of a notation that will work with any number of variables.

EXAMPLE 2.2.5. Express the following linear system as a matrix product.

$$\begin{aligned} x_1 + x_2 + x_3 &= 4 \\ 2x_1 + 2x_2 + 5x_3 &= 11 \\ 4x_1 + 6x_2 + 8x_3 &= 24 \end{aligned}$$

SOLUTION. Recall how we defined multiplication of a row vector and column vector at the beginning of this section. We use that as our inspiration. Define

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ 11 \\ 24 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 5 \\ 4 & 6 & 8 \end{bmatrix}$$

Of course, A is just the coefficient matrix of the system and b is the right hand side vector, which we have seen several times before. But now these take on a new significance. Notice that if we take the first row of A and multiply it by $\mathbf{x}$ we get the left hand side of the first equation of our system. Likewise for the second and third rows. Therefore, we may write in the language of matrices that

$$A\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 5 \\ 4 & 6 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 11 \\ 24 \end{bmatrix} = \mathbf{b}$$

Thus the system is represented very succinctly as $A\mathbf{x} = \mathbf{b}$. □

Once we understand this example, it is easy to see that the general abstract system that we examined in the first section of Chapter 1 can just as easily be abbreviated. Now we have a new way of looking at a system of equations: it is just like a simple first degree equation in one variable. Of course, the catch is that the symbols $A, \mathbf{x}, \mathbf{b}$ now represent an $m \times n$ matrix, $n \times 1$ and $m \times 1$ vectors, respectively. In spite of this the matrix multiplication idea is very appealing. For instance, it might inspire us to ask if we could somehow solve the system $A\mathbf{x} = \mathbf{b}$ by multiplying both sides of the equation by some kind of matrix "$1/A$" so as to cancel the A and get

$$(1/A)A\mathbf{x} = I\mathbf{x} = \mathbf{x} = (1/A)\mathbf{b}$$

We'll follow up on this idea in Section 2.5.

Here is another perspective on matrix-vector multiplication that gives a very powerful way of thinking about such multiplications. We will use this idea frequently in Chapter 3.

EXAMPLE 2.2.6. Interpret the matrix product of Example 2.2.5 as a linear combination of column vectors.

SOLUTION. Examine the system of this example and we see that the column $(1, 2, 4)$ appears to be multiplied by x_1. Similarly, the column $(1, 2, 6)$ is multiplied by x_2 and the column $(1, 5, 8)$ by x_3. Hence, if we use the same right hand side column $(4, 11, 24)$ as before, we obtain that this column can be expressed as a linear combination of column vectors, namely

$$x_1 \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 4 \\ 11 \\ 24 \end{bmatrix}$$

□

We could write the equation of the previous example very succinctly as follows: let A have columns $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, so that $A = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$ and let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$. Then

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3$$

Laws of Arithmetic

We have already seen that the laws of matrix arithmetic may not be quite the same as the ordinary arithmetic laws that we are use to. Nonetheless, as long as we don't assume a cancellation law or a commutative law for multiplication, things are pretty much what one might expect.

Laws of Matrix Multiplication. Let A, B, C be matrices of the appropriate sizes so that the following multiplications make sense, I a suitably sized identity matrix, and c and d scalars. Then

1. (Closure Law) AB is an $m \times n$ matrix.
2. (Associative Law) $(AB)C = A(BC)$
3. (Identity Law) $AI = A$ and $IB = B$
4. (Associative Law for Scalars) $c(AB) = (cA)B = A(cB)$
5. (Distributive Law) $(A + B)C = AC + BC$
6. (Distributive Law) $A(B + C) = AB + AC$

One can formally verify these laws by working through the definitions. For example, to verify the first half of the identity law, let $A = [a_{ij}]$ be an $m \times n$ matrix, so that $I = [\delta_{ij}]$ has to be I_n in order for the product AI to make sense. Now we see from the formal definition of matrix multiplication that

$$AI = [\sum_{k=1}^{n} a_{ik}\delta_{kj}] = [a_{ij} \cdot 1] = A$$

The middle equality follows from the fact that δ_{kj} is 0 unless $k = j$. Thus the sum collapses to a single term. A similar calculation verifies the other laws.

We end our discussion of matrix multiplication with a familiar looking notation that will prove to be extremely handy in the sequel.

NOTATION 2.2.7. Let A be a square $n \times n$ matrix and k a nonnegative integer. Then we define the *kth power* of A to be

$$A^k = \begin{cases} I_n & \text{if} \quad k = 0 \\ \underbrace{A \cdot A \cdot \ldots \cdot A}_{k \quad \text{times}} & \text{if} \quad k > 0 \end{cases}$$

As a simple consequence of this definition we have the standard exponent laws.

Laws of Exponents. For nonnegative integers i, j and square matrix A we have that

- $A^{i+j} = A^i \cdot A^j$
- $A^{ij} = (A^i)^j$

Notice that the law $(AB)^i = A^i B^i$ is missing. Why do you think it won't work with matrices?

2.2 Exercises

1. Express these systems of equations in the notation of matrix multiplication.

(a)
$$\begin{aligned} x_1 - 2x_2 + 4x_3 &= 3 \\ x_2 - x_3 &= 2 \\ -x_1 + 4x_3 &= 1 \end{aligned}$$

(b)
$$\begin{aligned} x - y - 3z &= 3 \\ 2x + 2y + 4z &= 10 \\ -x + z &= 3 \end{aligned}$$

(c)
$$\begin{aligned} x - 3y + 1 &= 0 \\ 2y &= 2 \\ -x + 3y &= 0 \end{aligned}$$

2. Let $A = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix}$ and verify the identity $(I + A + A^2)(I - A) = I - A^3$.

3. Express each system of Exercise 1 as an equation whose left hand side is a linear combination of the columns of the coefficient matrix (see Example 2.2.6).

4. Carry out these calculations or indicate they are impossible, given that $\mathbf{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 & 4 \end{bmatrix}$, and $C = \begin{bmatrix} 2 & 1+i \\ 0 & 1 \end{bmatrix}$

(a) $\mathbf{b}C\mathbf{a}$ (b) $\mathbf{ab}$ (c) $C\mathbf{b}$ (d) $(C\mathbf{b})\mathbf{a}$ (e) $C\mathbf{a}$ (f) $C(\mathbf{ab})$ (g) $\mathbf{ba}$ (h) $C(\mathbf{a}+\mathbf{b})$

5. Compute the product AB of the following pairs A, B, if possible.

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

(c) $\begin{bmatrix} 3 & 1 & 2 \\ 1 & 0 & 0 \\ 4 & 3 & 2 \end{bmatrix}, \begin{bmatrix} -5 & 4 & -2 \\ -2 & 3 & 1 \\ 1 & 0 & 4 \end{bmatrix}$ (d) $\begin{bmatrix} 3 & 1 \\ 1 & 0 \\ 4 & 3 \end{bmatrix}, \begin{bmatrix} -5 & 4 & -2 \\ -2 & 3 & 1 \end{bmatrix}$

(e) $\begin{bmatrix} -2 & 1 & -3 \end{bmatrix}, \begin{bmatrix} -5 & 4 & -2 \\ -2 & 3 & 1 \\ 1 & 0 & 4 \end{bmatrix}$ (f) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} -5 & 4 & -2 \end{bmatrix}$

6. Find examples of 2×2 matrices A and B that fulfill each of the following conditions.

(a) $(AB)^2 \neq A^2B^2$

(b) $AB \neq BA$

(c) $(A - B)^2 = A^2 - 2AB + B^2$

7. Prove that if two matrices A and B of the same size have the property that $A\mathbf{b} = B\mathbf{b}$ for every column vector $\mathbf{b}$ of the correct size for multiplication, then $A = B$. (Try using vectors with a single nonzero entry of 1.)

8. A square matrix A is said to be *nilpotent* if there is a positive integer m such that $A^m = 0$. Determine which of the following matrices are nilpotent; justify your answer.

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

9. A square matrix A is *idempotent* if $A^2 = A$. Determine which of the following matrices are idempotent.

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

10. Verify that the product $\mathbf{uv}^T$, where $\mathbf{u} = (1, 0, 2)$ and $\mathbf{v} = (-1, 1, 1)$, is a rank one matrix.

11. Verify that the associative law for scalars and both distributive laws hold for $c = 4$ and these matrices

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 2 \\ 0 & 3 \end{bmatrix} \quad C = \begin{bmatrix} 1+i & 1 \\ 1 & 2 \end{bmatrix}.$$

12. Prove that the associative law for scalars (page 60) is valid.

13. Prove that both distributive laws for matrices (page 60) are valid.

14. Let $A = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 & 0 \\ & 0 & 1 \\ \frac{5}{2} & -\frac{3}{2} & 0 \end{bmatrix}$ Compute $f(A)$ and $f(B)$ where $f(x) = 2x^3 + 3x - 5$. Here it is understood that when a square matrix is substituted for the variable x in a polynomial expression, the constant term has an implied coefficient of x^0 in it which becomes the identity matrix of appropriate size.

2.3. Applications of Matrix Arithmetic

We have already seen an important application of matrix multiplication to linear systems. We next examine a few more applications of the matrix multiplication idea which should reinforce the importance of this idea, as well as provide us with some interpretations of matrix multiplication.

Matrix Multiplication as Function Composition

The function idea is basic to mathematics. Recall that a *function* f is a rule of correspondence that assigns to each argument x in a set called its domain, a unique value $y = f(x)$ from a set called its range. Each branch of mathematics has its own special functions; for example, in calculus differentiable functions $f(x)$ are fundamental. Linear algebra also has its special functions. Suppose that $T(\mathbf{u})$ represents a function whose arguments $\mathbf{u}$ and values $\mathbf{v} = T(\mathbf{u})$ are vectors. We say the function T is *linear* if, for all vectors $\mathbf{u}, \mathbf{v}$ in the domain of T, and scalars c, d, we have

Linear Functions

$$T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v})$$

EXAMPLE 2.3.1. Show that the function T with domain the set of 2×1 vectors and definition by the formula

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = x$$

is a linear function.

SOLUTION. Let $\begin{bmatrix} x \\ y \end{bmatrix}$ and $\begin{bmatrix} z \\ w \end{bmatrix}$ be two elements in the domain of T and c, d any two scalars. Now compute

$$\begin{aligned} T\left(c\begin{bmatrix} x \\ y \end{bmatrix} + d\begin{bmatrix} z \\ w \end{bmatrix}\right) &= T\left(\begin{bmatrix} cx \\ cy \end{bmatrix} + \begin{bmatrix} dz \\ dw \end{bmatrix}\right) = T\left(\begin{bmatrix} cx+dz \\ cy+dw \end{bmatrix}\right) \\ &= cx + dz = cT\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) + dT\left(\begin{bmatrix} z \\ w \end{bmatrix}\right) \end{aligned}$$

Thus, T satisfies the definition of linear function. □

One can check that the function T just defined can be expressed as a matrix multiplication, namely, $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = [1,0]\begin{bmatrix} x \\ y \end{bmatrix}$. This kind of linear function gives yet another reason for defining matrix multiplication in the way that we do. More generally, let A be an $m \times n$ matrix and define a function T_A that maps $n \times 1$ vectors to $m \times 1$ vectors according to the formula

$$T_A(\mathbf{u}) = A\mathbf{u}.$$

First we verify that T is linear. Use the definition of T_A along with the distributive law of multiplication and associative law for scalars to obtain that

$$\begin{aligned} T_A(c\mathbf{u} + d\mathbf{v}) &= A(c\mathbf{u} + d\mathbf{v}) \\ &= A(c\mathbf{u}) + dA(\mathbf{v}) \\ &= c(A\mathbf{u}) + d(A\mathbf{v}) \\ &= cT_A(\mathbf{u}) + dT_A(\mathbf{v}) \end{aligned}$$

This proves that multiplication of vectors by a fixed matrix A is a linear function.

EXAMPLE 2.3.2. Use the associative law of matrix multiplication to show that the composition of matrix multiplication functions corresponds to the matrix product.

SOLUTION. For all vectors $\mathbf{u}$ and for suitably sized matrices A, B, we have by the associative law that $A(B\mathbf{u}) = (AB)\mathbf{u}$. In function terms, this means that $T_A(T_B(\mathbf{u})) = T_{AB}(\mathbf{u})$. Since this is true for all arguments $\mathbf{u}$, it follows that $T_A \circ T_B = T_{AB}$, which is what we were to show. □

We will have more to say about linear functions in Chapters 3 and 6, where they will go by the name of linear operators. For now we'll conclude our discussion of linear functions with an example that gives a hint of why the "linear" in "linear function."

EXAMPLE 2.3.3. Let L be the set of points (x, y) in the plane that satisfy the linear equation $y = x + 1$, $A = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$ and let $T_A(L) = \{T((x, y)) \mid (x, y) \in L\}$. Describe and sketch these sets in the plane.

SOLUTION. Of course the set L is just the straight line defined by the linear equation $y = x + 1$. To see what elements of $T_A(L)$ look like, we write a typical element of L in the form $(x, x + 1)$. Now calculate

$$T_A((x, x+1)) = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}\begin{bmatrix} x \\ x+1 \end{bmatrix} = \begin{bmatrix} 3x+1 \\ 6x+2 \end{bmatrix}.$$

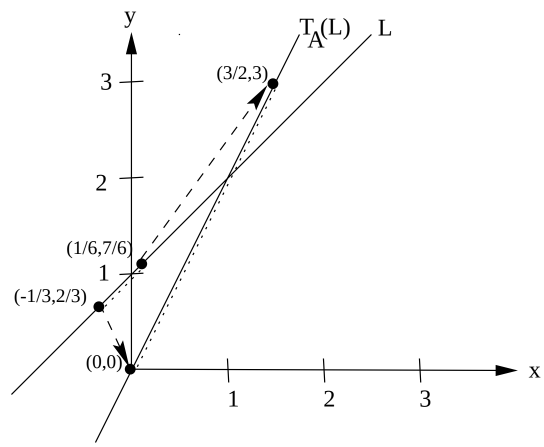


FIGURE 2.3.1. Action of T_A on line L given by $y = x + 1$ and points on L.

Now make the substitution $t = 3x + 1$ and we see that a typical element of $T_A(L)$ has the form $(t, 2t)$, where t is any real number, since x is arbitrary. We recognize these points as exactly the points on the line $y = 2x$. Thus, the function T_A maps the line $y = x + 1$ to the line $y = 2x$. Figure 2.3.1 illustrates this mapping as well as the fact that $T_A((-1/3, 2/3)) = (0, 0)$ and $T_A((1/6, 7/6)) = (3/2, 3)$.

Markov Chains as Matrix Products

A Markov chain is a certain type of matrix model which we will illustrate with an example.

EXAMPLE 2.3.4. Suppose two toothpaste companies compete for customers in a fixed market in which each customer uses either Brand A or Brand B. Suppose also that a market analysis shows that the buying habits of the customers fit the following pattern in the quarters that were analyzed: each quarter (three month period) 30% of A users will switch to B while the rest stay with A. Moreover, 40% of B users will switch to A in a given quarter, while the remaining B users will stay with B. If we *assume* that this pattern does not vary from quarter to quarter, we have an example of what is called a *Markov chain model.* Express the data of this model in matrix-vector language.

SOLUTION. Notice that if a_0 and b_0 are the fractions of the customers using A and B, respectively, in a given quarter, a_1 and b_1 the fractions of customers using A and B in the next quarter, then our hypotheses say that

$$\begin{aligned} a_1 &= 0.7a_0 + 0.4b_0 \\ b_1 &= 0.3a_0 + 0.6b_0 \end{aligned}$$

We could figure out what happens in the quarter after this by replacing the indices 1 and 0 by 2 and 1, respectively, in the preceding formula. In general, we replace the indices

$1, 0$ by $k, k+1$, to obtain

$$\begin{aligned} a_{k+1} &= 0.7a_k + 0.4b_k \\ b_{k+1} &= 0.3a_k + 0.6b_k \end{aligned}$$

We express this system in matrix form as follows: let $\mathbf{x}^{(k)} = \begin{bmatrix} a_k \\ b_k \end{bmatrix}$ and $A = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}$. Then the system may be expressed in the matrix form

$$\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)}$$

□

The state vectors $\mathbf{x}^{(k)}$ of the preceding example have the following property: each coordinate is non-negative and all the coordinates sum to 1. Such a vector is called a *probability distribution vector*. Also, the matrix A has the property that each of its columns is a probability distribution vector. Such a square matrix is called a transition matrix. In these terms we now give a precise definition of a Markov chain.

Markov Chain

DEFINITION 2.3.5. A *Markov chain* is a transition matrix A together with a probability distribution vector $\mathbf{x}^{(0)}$. The *state vectors* of this Markov chain are the vectors given by $\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)}$, $k = 0, 1, \ldots$.

Let us return to Example 2.3.4. The state vectors and transition matrices

$$\mathbf{x}^{(k)} = \begin{bmatrix} a_k \\ b_k \end{bmatrix} \qquad \text{and} \qquad A = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}$$

should play an important role. And indeed they do, for in light of our interpretation of a linear system as a matrix product, we see that the two equations of Example 2.3.4 can be written simply as $\mathbf{x}^{(1)} = Ax^{(0)}$. A little more calculation shows that

$$\mathbf{x}^{(2)} = A\mathbf{x}^{(1)} = A \cdot (A\mathbf{x}^{(0)}) = A^2\mathbf{x}^{(0)}$$

and in general,

$$\mathbf{x}^{(k)} = A\mathbf{x}^{(k-1)} = A^2\mathbf{x}^{(k-2)} = \ldots = A^k\mathbf{x}^{(0)}$$

Now we really have a very good handle on the Markov chain problem. Consider the following instance of our example.

EXAMPLE 2.3.6. In the notation of Example 2.3.4 suppose that initially Brand A has all the customers (i.e., Brand B is just entering the market). What are the market shares 2 quarters later? 20 quarters? Answer the same questions if initially Brand B has all the customers.

SOLUTION. To say that initially Brand A has all the customers is to say that the initial state vector is $\mathbf{x}^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Now do the arithmetic to find $\mathbf{x}^{(2)}$:

$$\begin{aligned} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} = \mathbf{x}^{(2)} = A^2\mathbf{x}^{(0)} &= \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} \left(\begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \\ &= \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} \begin{bmatrix} 0.7 \\ 0.3 \end{bmatrix} = \begin{bmatrix} .61 \\ .39 \end{bmatrix} \end{aligned}$$

Thus, Brand A will have 61% of the market and Brand B will have 39% of the market in the second quarter. We did not try to do the next calculation by hand, but rather used a computer to get the approximate answer:

$$\mathbf{x}^{(20)} = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}^{20} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} .57143 \\ .42857 \end{bmatrix}$$

Thus, after 20 quarters, Brand A's share will have fallen to about 57% of the market and Brand B's share will have risen to about 43%. Now consider what happens if the initial scenario is completely different, i.e., $\mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. We compute by hand to find that

$$\begin{aligned} \mathbf{x}^{(2)} &= \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} \left(\begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \\ &= \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} = \begin{bmatrix} .52 \\ .48 \end{bmatrix} \end{aligned}$$

Then we use a computer to find:

$$\mathbf{x}^{(20)} = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}^{20} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} .57143 \\ .42857 \end{bmatrix}$$

Surprise! We get the same answer as we did with a completely different initial condition. Is this just a coincidence? We will return to this example again in Chapter 3, where concepts introduced therein will cast new light on this model. □

Calculating Power of Graph Vertices

EXAMPLE 2.3.7. *(Dominance Directed Graphs)* Suppose you have incomplete data about four teams who have played each other, and further that the matches only have a winner and a loser, with no score attached. Given that the teams are identified by labels $1, 2, 3,$ and 4 we could describe a match by a pair of numbers (i, j) where we understand that this means that team i played and defeated team j (no ties allowed). Here is the given data:

$$\{(1, 2), (1, 4), (3, 1), (2, 3), (4, 2)\}$$

Give a reasonable graphical representation of this data.

SOLUTION. We can draw a picture of all the data that we are given by representing each team, as a point called a "vertex" and each match by connecting two points with an arrow, called a "directed edge", which points from the winner towards the loser in the match. See Figure 2.3.2 for the picture that we obtain. □

Consider the following question relating to Example 2.3.7. Given this incomplete data about the teams, how would we determine the ranking of each team in some reasonable way? In order to answer this question, we are going to introduce some concepts from graph theory which are useful modeling tools for many problems.

The data of Figure 2.3.2 is an example of a *directed graph*, a modeling tool which can be defined as follows. A *directed graph* (digraph for short) is a set V, whose elements are called *vertices*, together with a set or list (to allow for repeated edges) E of ordered Directed Graph

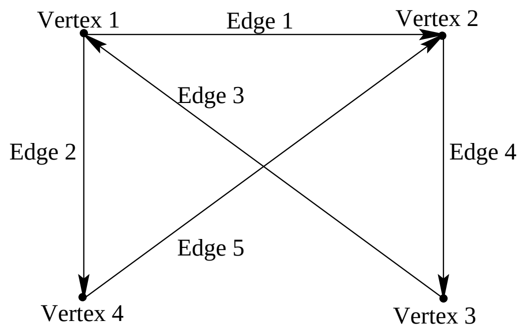


FIGURE 2.3.2. Data from Example 2.3.7

pairs with coordinates in V, whose elements are called *(directed) edges*. Another useful idea for us is the following: a *walk* in the digraph G is a sequence of digraph edges $(v_0, v_1), (v_1, v_2), ..., (v_{m-1}, v_m)$ which goes from vertex v_0 to vertex v_m. The *length* of the walk is m.

Here is an interpretation of "power" that has proved to be useful in many situations, including our own. The *power* of a vertex in a digraph is the number of walks of length 1 or 2 originating at the vertex. In our example, for instance, the power of vertex 1 is 4. Why only walks of length 1 or 2? For one thing, walks of length 3 introduce the possibility of *loops*, i.e., walks that "loop around" to the same point. It isn't very informative to find out that team 1 beat team 2 beat team 3 beat team 1. For another, information more than two hops away isn't very definitive. So we don't count it in the definition of power.

The type of digraph we are considering has no edges from a vertex to itself (so-called "self-loops") and for a pair of distinct vertices at most one edge connecting the two vertices. In other words, a team doesn't play itself and plays another team at most once. Such a digraph is called a *dominance-directed graph*. Although the notion of power of a point is defined for any digraph, it makes the most sense for dominance-directed graphs.

EXAMPLE 2.3.8. Find the power of each vertex in the graph of Example 2.3.7 and use this information to rank the teams.

SOLUTION. In this example we could find the power of all points by inspection of Figure 2.3.2. Let's do it: simple counting gives that the power of vertex 1 is 4, the power of vertex 3 is 3, and the power of vertices 2 and 4 is 2. Consequently, teams 2 and 4 are tied for last place, team 3 is in second place and team 1 is first. □

One can imagine situations (like describing the structure of the communications network pictured in Figure 2.3.3) where the edges shouldn't really have a direction, since connections are bidirectional. For such situations a more natural tool is the concept of a *graph*, which can be defined as follows: a *graph* is a set V, whose elements are called *vertices*, together with a set or list (to allow for repeated edges) E of *unordered* pairs with coordinates in V, called *edges*.

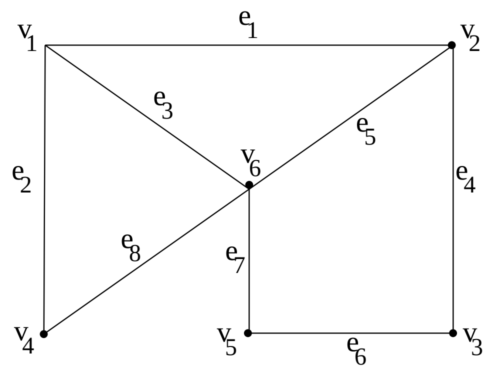


FIGURE 2.3.3. A communication network graph.

Just as with digraphs, we define A *walk* in the graph G is a sequence of digraph edges $(v_0, v_1), (v_1, v_2), ..., (v_{m-1}, v_m)$ which goes from vertex v_0 to vertex v_m. The *length* of the walk is m. For example, the graph of Figure 2.3.3 has vertex set $V = \{v_1, v_2, v_3, v_4, v_5, v_6\}$ and edge set $E = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\}$, with $e_1 = (v_1, v_2)$, etc, as in the figure. Also, the sequence e_1, e_4, e_7 is a walk from vertex v_1 to v_5 of length 2. Just as with digraphs, we can define the power of a vertex in any graph as the number of walks of length at most 2 originating at the vertex.

A very sensible question to ask about these examples: how could we write a computer program to compute powers? More generally, how can we compute the total number of walks of a certain length? Here is a key to the answer: all the information about our graph (or digraph) can be stored in its *adjacency matrix*. In general, this is defined to be a square matrix whose rows and columns are indexed by the vertices of the graph and whose (i, j)th entry is the number of edges going from vertex i to vertex j (it is 0 if there are none). Here we understand that a directed edge of a digraph must start at i and end at j, while no such restriction applies to the edges of a graph.

Just for the record, if we designate the adjacency matrix of the digraph of Figure 2.3.2 by A and the adjacency matrix of the graph of Figure 2.3.3 by B, then

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Notice that we could reconstruct the entire digraph or graph from this matrix. Also notice that in the adjacency matrix for a graph, an edge gets accounted for twice, since it can be thought of as proceeding from one vertex to the other, or from the other to the one.

For a general graph with n vertices and adjacency matrix $A = [a_{ij}]$, we can use this matrix to compute powers of vertices without seeing a picture of the graph. To count up the walks of length 1 emanating from vertex i: simply add up the elements of the ith row of A. Now what about the paths of length 2? Observe that there is an edge from i to k and then from k to j precisely when the product $a_{ik}a_{kj}$ is equal to 1. Otherwise, one of the factors will be 0 and therefore the product is 0. So the number of paths of length 2 from vertex i to vertex j is the familiar sum

$$a_{i1}a_{1j} + a_{i2}a_{2j} + \cdots + a_{in}a_{nj}$$

This is just the (i, j)th entry of the matrix A^2. A similar argument shows the following fact:

THEOREM 2.3.9. *If A is the adjacency matrix of the graph G, then the (i, j)th entry of A^r gives the number of walks of length r starting at vertex i and ending at vertex j.*

Since the power of vertex i is the number of all paths of length 1 or 2 emanating from vertex i, we have the following key fact:

THEOREM 2.3.10. *If A is the adjacency matrix of the digraph G, then the power of the ith vertex is the sum of all entries in the ith row of the matrix $A + A^2$.*

EXAMPLE 2.3.11. Use the preceding facts to calculate the powers of all the vertices in the digraph of Example 2.3.7.

SOLUTION. Using the matrix A above we calculate that

$$A + A^2 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

An easy way to sum each row is to multiply $A + A^2$ on the right by a column of 1's, but in this case we see immediately that the power of vertex 1 is 4, the power of vertex 3 is 3, and the power of vertices 2 and 4 is2, which is consistent with what we observed earlier by inspection of the graph. □

Difference Equations

The idea of a difference equation has numerous applications in mathematics and computer science. In the latter field, these equations often go by the name of "recurrence relations." They can be used for a variety of applications ranging from population modeling to analysis of complexity of algorithms. We will introduce them by way of a simple financial model.

EXAMPLE 2.3.12. Suppose that you invest in a contractual fund that has the stipulation that you must invest in the funds for three years before you can receive any return on your investment (with a positive first year investment). Thereafter, you are vested in the fund and may remove your money at any time. While you are vested in the fund, annual returns are calculated as follows: money that was in the fund one year ago earns nothing, while money that was in the fund two years ago earns 6% of its value and money that was in the fund three years ago earns 12% of its value. Find an equations that describes your investment's growth.

SOLUTION. Let a_k be the amount of your investment in the kth year. The numbers a_0, a_1, a_2 represent your investments for the first three years (we're counting from 0.) Consider the 3rd year amount a_3. According to your contract, your total funds in the 3rd year will be

$$a_3 = a_2 + 0.06a_1 + 0.12a_0$$

Now it's easy to write out a general formula for a_{k+3} in terms of the preceding three terms, using the same line of thought, namely

$$a_{k+3} = a_{k+2} + 0.06a_{k+1} + 0.12a_k, \quad k = 0, 1, 2, \ldots \tag{2.3.1}$$

This is the desired formula. □

In general, a *homogeneous linear difference equation* (or *recurrence relation*) of order m in the variables $a_0, a_1, \ldots$ is an equation of the form

$$a_{k+m} + c_{m-1}a_{k+m-1} + \ldots + c_1a_{k+1} + c_0a_k = 0, \quad k = 0, 1, 2, \ldots$$

Notice that such an equation cannot determine the numbers $a_0, a_1, \ldots, a_{k-1}$. These values have to be initially specified, just as in our fund example. Notice that in our fund example, we have to bring all terms of Equation 2.3.1 to the left hand side to obtain the difference equation form

$$a_{k+3} - a_{k+2} - 0.06a_{k+1} - 0.12a_k = 0$$

Now we see that $c_2 = -1$, $c_1 = -0.06$, and $c_0 = -0.12$.

There are many ways to solve difference equations; we are not going to give a complete solution to this problem at this point – we postpone this issue to Chapter 5, where we introduce eigenvalues and eigenvectors. However, we can now show how to turn a difference equation as given above into a matrix equation. We'll illustrate the key idea with our fund example. The secret is to identify the right vector variables. To this end, define an indexed vector $\mathbf{x}_k$ by the formula

$$\mathbf{x}_k = \begin{bmatrix} a_{k+2} \\ a_{k+1} \\ a_k \end{bmatrix}, \quad k = 0, 1, 2, \ldots$$

We see that

$$\mathbf{x}_{k+1} = \begin{bmatrix} a_{k+3} \\ a_{k+2} \\ a_{k+1} \end{bmatrix}$$

from which it is easy to check that since $a_{k+3} = a_{k+2} + 0.06a_{k+1} + 0.12a_k$, we have

$$\mathbf{x}_{k+1} = \begin{bmatrix} 1 & 0.06 & 0.12 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x}_k = A\mathbf{x}_k$$

This is the matrix form we seek. Notice that it seems to have a lot in common with the Markov chains examined earlier in this section, in that we pass from one "state vector" to another by multiplication by a fixed "transition matrix" A.

2.3 Exercises

1. Let a function of the vector variable $\mathbf{x} = (x_1, x_2)$ be given by the formula

$$T(\mathbf{x}) = (x_1 + x_2, 2x_1, 4x_2 - x_1)$$

Show how to express this function as a matrix multiplication and deduce that it is a linear function.

2. Prove that if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is a real 2×2 matrix, then the matrix multiplication function maps a line through the origin onto a line through the origin or a point. *Hint:* Recall that points on a non-vertical line through the origin have the form (x, mx).

3. Determine the effect of the matrix multiplication function T_A on the x-axis, y-axis, and the points $(\pm 1, \pm 1)$, where A is one of the following, and sketch your results.

(a) $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$ (c) $\frac{1}{5}\begin{bmatrix} -3 & -4 \\ -4 & 3 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ (f) $\frac{1}{5}\begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}$

4. Use the definition of matrix multiplication function to show that if $T_A = T_B$, then $A = B$. (See Exercise 7 of Section 2.)

5. Inpection of the graph in Figure 2.3.1 of the matrix multiplication function T_A from Example 2.3.3 suggests that this function has a *fixed point*, that is, a vector (x, y) such that $T_A((x, y)) = (x, y)$. Describe this point on the graph and calculate it algebraically.

6. Suppose a Markov chain has transition matrix $\begin{bmatrix} .1 & .3 & 0 \\ 0 & .4 & 1 \\ .9 & .3 & 0 \end{bmatrix}$ and initial state $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.
(a) Calculate the second state vector of the system. (b) Use a computer to decide experimentally whether or not the system tends to a limiting steady state vector. If so, what is it?

7. You are given that a Markov chain has transition matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$,where all entries are between 0 and 1. Show how this matrix can be described using only two variables.

8. You are given that the digraph G has vertex set $V = \{1, 2, 3, 4, 5\}$ and edge set $E = \{(2, 1), (1, 5), (2, 5), (5, 4), (4, 2), (4, 3), (3, 2)\}$. Do the following for the graph G.

(a) Sketch a picture of the graph.

(b) Find the adjacency matrix of the graph.

(c) Find the power of each vertex of the graph. Which vertices are the strongest?

(d) Is this graph a dominance-directed graph?

9. A digraph has the following adjacency matrix:

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

(a) Draw a picture of this digraph (this graph has some self-loops).

(b) Compute the power of each node and the total number of walks in the digraph of length at most 5. Use a computer algebra system or calculator to help you with the computations.

10. Find the adjacency matrix of the graph of Figure 2.3.3 and use it to determine the total number of walks of length of length less than or equal to 3 starting at the node v_6.

11. Convert the fourth order difference equation

$$a_{k+4} - 2a_{k+3} + 3a_{k+2} - 4a_{k+1} + 5a_k = 0$$

into vector form.

12. Suppose that in Example 2.3.12 you invest \$1,000 initially (the zeroth year) and no further amounts. Make a table of the value of your investment for years 0 to 12. Also include a column which calculates the annual interest rate that your investment is earning each year, based on the current and previous years' value. What conclusions do you draw? You will need a computer or calculator for this exercise.

2.4. Special Matrices and Transposes

There are certain types of matrices that are so important that they have acquired names of their own. We are going to introduce some of these in this section, as well as one more matrix operation which has proved to be a very practical tool in matrix analysis, namely the operation of transposing a matrix.

Elementary Matrices and Gaussian Elimination

We are going to show a new way to execute the elementary row operations used in Gaussian elimination. Recall the shorthand we used:

- E_{ij} : The elementary operation of *switching the ith and jth rows* of the matrix.
- $E_i(c)$: The elementary operation of *multiplying the ith row by the nonzero constant c*.

- $E_{ij}(d)$: The elementary operation of *adding d times the jth row to the ith row.*

From now on we will use the very same symbols to represent matrices. The size of the matrix will depend on the context of our discussion, so the notation is ambiguous, but it is still very useful.

NOTATION 2.4.1. An *elementary matrix* of size n is obtained by performing the corresponding elementary row operation on the identity matrix I_n. We denote the resulting matrix by the same symbol as the corresponding row operation.

EXAMPLE 2.4.2. Describe the following elementary matrices of size $n = 3$:

$$(a)\ E_{13}(-4)\,, \quad (b)\ E_{21}(3), \quad (c)\ \ E_{23}\,, \quad (d)\ \ E_1(1/2)$$

SOLUTION. We start with

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

For part (a) we add -4 times the 3rd row of I_3 to its first row to obtain

$$E_{13}(-4) = \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

For part (b) add 3 times the first row of I_3 to its second row to obtain

$$E_{21}(3) = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

For part (c) interchange the second and third rows of I_3 to obtain that

$$E_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Finally, for part (d) we multiply the first row of I_3 by $1/2$ to obtain

$$E_1(1/2) = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

□

What good are these matrices? One can see that the following fact is true:

THEOREM 2.4.3. *Let $C = BA$ be a product of two matrices and perform an elementary row operation on C. Then the same result is obtained if one performs the same elementary operation on the matrix B and multiplies the result by A on the right.*

We won't give a formal proof of this statement, but it isn't hard to see why it is true. For example, suppose one interchanges two rows, say the ith and jth, of $C = BA$ to obtain a new matrix D. How do we get the ith or jth row of C? Answer: multiply the corresponding row of B by the matrix A. Therefore, we would obtain D by interchanging the ith and jth rows of B and multiplying the result by the matrix A, which is exactly what the Theorem says. Similar arguments apply to the other elementary operations.

Now take $B = I$, and we see from the definition of elementary matrix and Theorem 2.4.3 that the following is true.

Elementary Operations as Matrix Multiplication

COROLLARY 2.4.4. *If an elementary row operation is performed on a matrix A to obtain a matrix A', then $A' = EA$, where E is the elementary matrix corresponding to the elementary row operation performed.*

The meaning of this corollary is that we accomplish an elementary row operation by multiplying by the corresponding elementary matrix on the left. Of course, we don't need elementary matrices to accomplish row operations; but they give us another perspective on row operations.

EXAMPLE 2.4.5. Express these calculations of Example 1.3.1 of Chapter 1 in matrix product form:

$$\begin{bmatrix} 2 & -1 & 1 \\ 4 & 4 & 20 \end{bmatrix} \xrightarrow[E_{12}]{} \begin{bmatrix} 4 & 4 & 20 \\ 2 & -1 & 1 \end{bmatrix}$$

$$\xrightarrow[E_1(1/4)]{} \begin{bmatrix} 1 & 1 & 5 \\ 2 & -1 & 1 \end{bmatrix} \xrightarrow[E_{21}(-2)]{} \begin{bmatrix} 1 & 1 & 5 \\ 0 & -3 & -9 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 5 \\ 0 & -3 & -9 \end{bmatrix} \xrightarrow[E_2(-1/3)]{} \begin{bmatrix} 1 & 1 & 5 \\ 0 & 1 & 3 \end{bmatrix} \xrightarrow[E_{12}(-1)]{} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix}$$

SOLUTION. One point to be careful about: the order of elementary operations. We compose the elementary matrices on the left in that same order that the operations are done. Thus we may state the above calculations in the concise form

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix} = E_{12}(-1)E_2(-1/3)E_{21}(-2)E_1(1/4)E_{12} \begin{bmatrix} 2 & -1 & 1 \\ 4 & 4 & 20 \end{bmatrix}$$

□

It is important to read this line carefully and understand how it follows from the long form above. This conversion of row operations to matrix multiplication will prove to be very practical in the next section.

Some Matrices with Simple Structure

Certain types of matrices have already turned up frequently in our discussions. For example, the identity matrices are particularly easy to deal with. Another example is the reduced row echelon form. So let us classify some simple matrices and attach names to them. The simplest conceivable matrices are zero matrices. We have already seen that they are important in matrix addition arithmetic. What's next? For square matrices, we have the following definitions, in ascending order of complexity.

DEFINITION 2.4.6. Let $A = [a_{ij}]$ be a square $n \times n$ matrix. Then A is

- *Scalar* if $a_{ij} = 0$ and $a_{ii} = a_{jj}$ for all $i \neq j$. (Equivalently: $A = cI_n$ for some scalar c, which explains the term "scalar.")

- *Diagonal* if $a_{ij} = 0$ for all $i \neq j$. (Equivalently: the off-diagonal entries of A are 0.)
- *(Upper) triangular* if $a_{ij} = 0$ for all $i > j$. (Equivalently: the sub-diagonal entries of A are 0.)
- *(Lower) triangular* if $a_{ij} = 0$ for all $i < j$. (Equivalently: the super-diagonal entries of A are 0.)
- *Triangular* if the matrix is upper or lower triangular.
- *Strictly triangular* if it is triangular and the diagonal entries are also zero.

SOLUTION. The index conditions that we use above have simple interpretations. For example, the entry a_{ij} with $i > j$ is located further down than over, since the row number is larger than the column number. Hence, it resides in the "lower triangle" of the matrix. Similarly, the entry a_{ij} with $i < j$ resides in the "upper triangle." Entries a_{ij} with $i = j$ reside along the main diagonal of the matrix. See the adjacent figure for a picture of these triangular regions of the matrix.

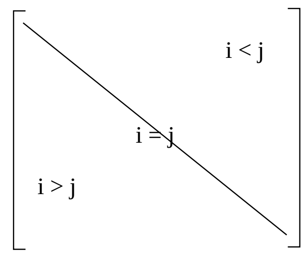


Figure 2.4.1: Matrix regions

EXAMPLE 2.4.7. Classify the following matrices (elementary matrices are understood to be 3×3) in the terminology of Definition 2.4.6.

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

$$\text{(d)} \begin{bmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ 3 & 2 & 2 \end{bmatrix} \quad \text{(e)} \begin{bmatrix} 0 & 2 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{(f) } E_{21}(3) \quad \text{(g) } E_2(-3)$$

SOLUTION. Notice that (a) is not scalar, since diagonal entries differ from each other, but it is a diagonal matrix, since the off-diagonal entries are all 0. On the other hand, the matrix of (b) is really just $2I_3$, so this matrix is a scalar matrix. Matrix (c) has all terms below the main diagonal equal to 0, so this matrix is triangular and, specifically, upper triangular. Similarly, matrix (d) is lower triangular. Matrix (e) is clearly upper triangular, but it is also strictly upper triangular since the diagonal terms themselves are 0. Finally, we have

$$E_{21}(3) = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad E_2(-3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

so that $E_{21}(3)$ is (lower) triangular and $E_2(-3)$ is a diagonal matrix. $\square$

Here is another kind of structured matrix that occurs frequently enough in applications to warrant a name. In Example 1.1.5 of Chapter 1 we saw that an approximation to a certain diffusion problem led to matrices of the form

$$A_5 = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix}$$

If we want more accurate solutions to the original problem, we would need to solve systems with coefficient matrix A_n, where n is larger than 5. Notice that the only nonzero entries of such a matrix are those along the main diagonal, the entries along the first subdiagonal and first superdiagonal. Such a matrix is called a *tridiagonal matrix*. Formally speaking, these are the square matrices $A = [a_{ij}]$ such that $a_{ij} = 0$ if $i > j + 1$ or $j > i + 1$.

Block matrices

Another type of matrix that occurs frequently enough to be discussed is a *block matrix*. Actually, we already used the idea of blocks when we describe the augmented matrix of the system $A\mathbf{x} = \mathbf{b}$ as the matrix $\tilde{A} = [A \,|\, \mathbf{b}]$. We say that $\tilde{A}$ has the *block*, or *partitioned*, form $[A, \mathbf{b}]$. What we are really doing is partitioning the matrix $\tilde{A}$ by inserting a vertical line between elements. There is no reason we couldn't partition by inserting more vertical lines or horizontal lines as well, and this partitioning leads to the blocks. The main point to bear in mind when using the block notation is that the blocks must be correctly sized so that the resulting matrix makes sense. The main virtue of the block form that results from partitioning is that for purposes of matrix addition or multiplication, we can treat the blocks rather like scalars, provided the addition or multiplication that results makes sense. We will use this idea from time to time without fanfare. One could go through a formal description of partitioning and proofs; we won't. Rather, we'll show how this idea can be used by example.

EXAMPLE 2.4.8. Use block multiplication to simplify the following multiplication

SOLUTION. Here is the blocking that we want to use. It makes the column numbers of the blocks on the left match the row numbers of the blocks on the right:

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

We see that these submatrices are built from zero matrices and these blocks:

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

Now we can work this product out by interpreting it as

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}\begin{bmatrix} 0 & C \\ 0 & I_2 \end{bmatrix} = \begin{bmatrix} A \cdot 0 + 0 \cdot 0 & A \cdot C + 0 \cdot I_2 \\ 0 \cdot 0 + B \cdot 0 & 0 \cdot C + B \cdot I_2 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 & 4 & 3 \\ 0 & 0 & 10 & 7 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

□

For another example of block arithmetic, examine Example 2.2.6 and the discussion following it. There we view a matrix as blocked into its respective columns, and a column vector as blocked into its rows, to obtain

$$A\mathbf{x} = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \mathbf{a}_3 x_3$$

Transpose of a Matrix

Sometimes we would prefer to work with a different form of a given matrix that contains the same information. Transposes are operations that allow us to do that. The idea of transposing is simple: interchange rows and columns. It turns out that for complex matrices, there is an analogue which is not quite the same thing as transposing, though it yields the same result when applied to real matrices. This analogue is called the Hermitian transpose. Here are the appropriate definitions.

DEFINITION 2.4.9. Let $A = [a_{ij}]$ be an $m \times n$ matrix with (possibly) complex entries. Then the *transpose* of A is the $n \times m$ matrix A^T obtained by interchanging the rows and columns of A, so that the (i, j)th entry of A^T is a_{ji}. The *conjugate* of A is the matrix $\overline{A} = [\overline{a_{ij}}]$. Finally, the *Hermitian transpose* of A is the matrix $A^H = \overline{A}^T$.

Notice that in the case of a real matrix (that is, a matrix with real entries) A there is no difference between transpose and Hermitian transpose, since in this case $A = \overline{A}$. Consider these examples.

EXAMPLE 2.4.10. Compute the transpose and Hermitian transpose of the following matrices:

$$\text{(a)} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}, \quad \text{(b)} \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}, \quad \text{(c)} \begin{bmatrix} 1 & 1+i \\ 0 & 2i \end{bmatrix}$$

SOLUTION. For matrix (a) we have

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}^H = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix}$$

Notice, by the way how the dimensions of a transpose get switched from the original.

For matrix (b) we have

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}^H = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}^T = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}$$

and for matrix (c) we have

$$\begin{bmatrix} 1 & 1+i \\ 0 & 2i \end{bmatrix}^H = \begin{bmatrix} 1 & 0 \\ 1-i & -2i \end{bmatrix}, \quad \begin{bmatrix} 1 & 1+i \\ 0 & 2i \end{bmatrix}^T = \begin{bmatrix} 1 & 0 \\ 1+i & 2i \end{bmatrix}$$

In this case, transpose and Hermitian transpose are not the same. □

Even when dealing with vectors alone, the transpose notation is rather handy. For example, there is a bit of terminology that comes from tensor analysis (a branch of higher linear algebra used in many fields including differential geometry, engineering mechanics and relativity) that can be expressed very concisely with transposes:

Inner and Outer Products

DEFINITION 2.4.11. Let u and v be column vectors of the same size, say $n \times 1$. Then the *inner product* of $\mathbf{u}$ and $\mathbf{v}$ is the scalar quantity $\mathbf{u}^T\mathbf{v}$ and the *outer product* of $\mathbf{u}$ and $\mathbf{v}$ is the $n \times n$ matrix $\mathbf{u}\mathbf{v}^T$.

EXAMPLE 2.4.12. Compute the inner and outer products of the vectors $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}$.

SOLUTION. Here we have the inner product

$$\mathbf{u}^T\mathbf{v} = [2, -1, 1] \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} = 2 \cdot 3 + (-1)4 + 1 \cdot 1 = 3$$

while the outer product is

$$\mathbf{u}\mathbf{v}^T = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} [3, 4, 1] = \begin{bmatrix} 2 \cdot 3 & 2 \cdot 4 & 2 \cdot 1 \\ -1 \cdot 3 & -1 \cdot 4 & -1 \cdot 1 \\ 1 \cdot 3 & 1 \cdot 4 & 1 \cdot 1 \end{bmatrix} = \begin{bmatrix} 6 & 8 & 2 \\ -3 & -4 & -1 \\ 3 & 4 & 1 \end{bmatrix}$$

□

Here are a few basic laws relating transposes to other matrix arithmetic that we have learned. These laws remain correct if transpose is replaced by Hermitian transpose, with one exception: $(cA)^H = \overline{c}A^H$.

Laws of Matrix Transpose. Let A and B be matrices of the appropriate sizes so that the following operations make sense, and c a scalar. Then

1. $(A + B)^T = A^T + B^T$
2. $(AB)^T = B^T A^T$
3. $(cA)^T = cA^T$
4. $(A^T)^T = A$

These laws are easily verified directly from definition. For example, if $A = [a_{ij}]$ and $B = [b_{ij}]$ are $m \times n$ matrices, then we have that $(A + B)^T$ is the $n \times m$ matrix given by

$$\begin{aligned}(A+B)^T &= [a_{ij}+b_{ij}]^T = [a_{ji}+b_{ji}] \\ &= [a_{ji}] + [b_{ji}] \\ &= A^T + B^T\end{aligned}$$

The other laws are proved similarly.

We will require explicit formulas for transposes of the elementary matrices in some later calculations. Notice that the matrix $E_{ij}(c)$ is a matrix with 1's on the diagonal and 0's elsewhere, except that the (i,j)th entry is c. Therefore, transposing switches the entry c to the (j,i)th position and leaves all other entries unchanged. Hence $E_{ij}(c)^T = E_{ji}(c)$. With similar calculations we have these facts

Transposes of Elementary Matrices

- $E_{ij}^T = E_{ij}$
- $E_i(c)^T = E_i(c)$
- $E_{ij}(c)^T = E_{ji}(c)$

Elementary Column Operations and Matrices

These formulas have an interesting application. Up to this point we have only considered elementary row operations. However, there are situations in which elementary *column* operations on the columns of a matrix are useful. If we want to use such operations, do we have to start over, reinvent elementary column matrices, and so forth? The answer is no and the following example gives an indication of why the transpose idea is useful. This example shows how to do column operations in the language of matrix arithmetic. In a nutshell, here's the basic idea: suppose we want to do an elementary column operation on a matrix A corresponding to elementary row operation E to get a new matrix B from A. To do this, turn the columns of A into rows, do the row operation and then transpose the result back to get the matrix B that we want. In algebraic terms:

$$B = (EA^T)^T = (A^T)^T E^T = AE^T$$

So all we have to do to perform an elementary column operation is multiply by the transpose of the corresponding elementary row matrix on the right. Thus we see that the transposes of elementary row matrices could reasonably be called elementary column matrices.

EXAMPLE 2.4.13. Let A be a given matrix. Suppose that we wish to express the result B of swapping the second and third columns of A, followed by adding -2 times the first column to the second, as a product of matrices. How can this be done? Illustrate the procedure with the matrix.

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

SOLUTION. Apply the preceding remark twice to obtain that

$$B = AE_{23}^T E_{21}(-2)^T = AE_{23}E_{12}(-2).$$

Thus we have

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

as a matrix product. □

A very important type of special matrix is one which is invariant under the operation of transposing. These matrices turn up naturally in applied mathematics. They have some very remarkable properties that we will study in Chapters 4, 5 and 6.

DEFINITION 2.4.14. The matrix A is said to be *symmetric* if $A^T = A$ and *Hermitian* if $A^H = A$. (Equivalently, $a_{ij} = a_{ji}$ and $a_{ij} = \overline{a_{ji}}$, for all i, j, respectively.)

From the laws of transposing elementary matrices above we see right away that E_{ij} and $E_i(c)$ supply us with examples of symmetric matrices. Here are a few more.

EXAMPLE 2.4.15. Are the following matrices symmetric or Hermitian?

$$\text{(a)} \begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix}, \quad \text{(b)} \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad \text{(c)} \begin{bmatrix} 1 & 1+i \\ 1+i & 2i \end{bmatrix}$$

SOLUTION. For matrix (a) we have

$$\begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix}^H = \begin{bmatrix} 1 & \overline{1+i} \\ \overline{1-i} & 2 \end{bmatrix}^T = \begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix}$$

Hence this matrix is Hermitian. However, it is *not* symmetric since the $(1, 2)$th and $(2, 1)$th entries differ. Matrix (b) is easily seen to be symmetric by inspection. Matrix (c) is symmetric since the $(1, 2)$th and $(2, 1)$th entries agree, but it is not Hermitian since

$$\begin{bmatrix} 1 & 1+i \\ 1-i & 2i \end{bmatrix}^H = \begin{bmatrix} 1 & \overline{1+i} \\ \overline{1-i} & \overline{2i} \end{bmatrix}^T = \begin{bmatrix} 1 & 1+i \\ 1-i & -2i \end{bmatrix}$$

and this last matrix is clearly not equal to matrix (c). □

EXAMPLE 2.4.16. Consider the quadratic form

$$Q(x, y, z) = x^2 + 2y^2 + z^2 + 2xy + yz + 3xz.$$

Express this function in terms of matrix products and transposes.

SOLUTION. Write the quadratic form as

$$\begin{aligned} x(x + 2y + 3z) + y(2y + z) + z^2 &= \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} x + 2y + 3z \\ 2y + z \\ z \end{bmatrix} \\ &= \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\ &= \mathbf{x}^T A \mathbf{x}, \end{aligned}$$

where $\mathbf{x} = (\mathbf{x}, \mathbf{y}, \mathbf{z})$ and $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. □

Rank of the Matrix Transpose

An important question to ask is how the rank of a matrix transpose (or Hermitian transpose) is connected to the rank of matrix. We focus on transposes. First we need the following

THEOREM 2.4.17. *Let A, B be matrices such that the product AB is defined. Then*

$$\operatorname{rank} AB \le \operatorname{rank} A$$

PROOF. Let E be a product of elementary matrices such that $EA = R$, where R is the reduced row echelon form of A. If $\operatorname{rank} A = r$, then the first r rows of A have leading entries of 1, while the remaining rows are zero rows. Also, we saw in Chapter 1 that elementary row operations do not change the rank of a matrix since, according to Corollary 1.4.9 they do not change the reduced row echelon form of a matrix. Therefore,

$$\operatorname{rank} AB = \operatorname{rank} E(AB) = \operatorname{rank}(EA)B = \operatorname{rank} RB$$

Now the matrix RB has the same number of rows as R and the first r of these rows may or may not be nonzero, but the remaining rows must be zero rows, since they result from multiplying columns of B by the zero rows of R. If we perform elementary row operations to reduce RB to its reduced row echelon form we will possibly introduce more zero rows than R has. Consequently, $\operatorname{rank} RB \le r = \operatorname{rank} A$, which completes the proof. □

THEOREM 2.4.18. *For any matrix A,*

$$\operatorname{rank} A = \operatorname{rank} A^T$$

PROOF. As in the previous theorem, let E be a product of elementary matrices such that $EA = R$, where R is the reduced row echelon form of A. If $\operatorname{rank} A = r$, then the first r rows of R have leading entries of 1 whose column numbers form an increasing sequence, while the remaining rows are zero rows. Therefore, $R^T = A^T E^T$ is a matrix whose columns have leading entries of 1 and whose row numbers form an increasing sequence. Use elementary row operations to clear out the nonzero entries below each column with a leading 1 to obtain a matrix whose rank is equal to the number of such leading entries, i.e., equal to r. Thus, $\operatorname{rank} R^T = r$.

From Theorem 2.4.17 we have that $\operatorname{rank} A^T E^T \le \operatorname{rank} A^T$. It follows that

$$\operatorname{rank} A = \operatorname{rank} R^T = \operatorname{rank} A^T E^T \le \operatorname{rank} A^T$$

If we substitute the matrix A^T for the matrix A in this inequality, we obtain that

$$\operatorname{rank} A^T \le \operatorname{rank}(A^T)^T = \operatorname{rank} A$$

It follows from these two inequalities that $\operatorname{rank} A = \operatorname{rank} A^T$, which is what we wanted to show. □

It is instructive to see how a specific example might work out in the preceding proof. For example, R might look like this, where an "x" designates an arbitrary entry,

$$R = \begin{bmatrix} 1 & 0 & x & 0 & x \\ 0 & 1 & x & 0 & x \\ 0 & 0 & 0 & 1 & x \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

so that R^T would look like this

$$R^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ x & x & 0 & 0 \\ 0 & 0 & 1 & 0 \\ x & x & x & 0 \end{bmatrix}$$

Thus if we use elementary row operations to zero out the entries below a column pivot, all entries to the right and below this pivot are unaffected by these operations. Now start with the leftmost column and proceed to the right, zeroing out all entries under each column pivot. The result is a matrix that looks like

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

Now swap rows to move the zero rows to the bottom if necessary and we see that the reduced row echelon form of R^T has exactly as many nonzero rows as did R, that is, r nonzero rows.

A first application of this important fact is to give a fuller picture of the rank of a product of matrices than Theorem 2.4.17:

COROLLARY 2.4.19. *If the product AB is defined, then*

Rank of Matrix Product

$$\operatorname{rank} AB \leq \min\{\operatorname{rank} A, \operatorname{rank} B\}$$

PROOF. We know from Theorem 2.4.17 that

$$\operatorname{rank} AB \leq \operatorname{rank} A \text{ and } \operatorname{rank} B^T A^T \leq \operatorname{rank} B^T$$

Since $B^T A^T = (AB)^T$, Theorem 2.4.18 tells us that

$$\operatorname{rank} B^T A^T = \operatorname{rank} AB \text{ and } \operatorname{rank} B^T = \operatorname{rank} B$$

Put all this together and we have

$$\operatorname{rank} AB = \operatorname{rank} B^T A^T \leq \operatorname{rank} B^T = \operatorname{rank} B$$

It follows that $\operatorname{rank} AB$ is at most the smaller of $\operatorname{rank} A$ and $\operatorname{rank} B$, which is what the corollary asserts. □

2.4 Exercises

1. Write out explicitly what the following 4×4 elementary matrices are:

(a) $E_{24}(3)$ (b) E_{14} (c) $E_3(2)$ (d) $E_{23}^T(-1)$ (e) E_{24}^T

2. Describe the effects of these multiplications as column operations on the matrix A.

(a) AE_{12} (b) $AE_{13}(-2)$ (c) $AE_2(-1)E_{14}(3)$ (d) $AE_{14}E_{41}$

3. For each of the following matrices, identify all of the simple structure descriptions that apply to the matrix.

(a) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 0 \\ 3 & 1 \end{bmatrix}$ (c) I_3 (d) $\begin{bmatrix} 2 & 1 & 4 & 2 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

4. Calculate the matrix product AB using block multiplication, where

$$A = \begin{bmatrix} R & 0 \\ S & T \end{bmatrix} \qquad B = \begin{bmatrix} U \\ V \end{bmatrix} \qquad R = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}$$

$$S = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \qquad T = \begin{bmatrix} 1 & -1 \\ 2 & 2 \end{bmatrix} \qquad U = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} \qquad V = \begin{bmatrix} 3 & 1 \\ 0 & 1 \end{bmatrix}$$

5. Let A and B be square matrices and suppose that the matrix $M = \begin{bmatrix} A & C \\ 0 & D \end{bmatrix}$ in block form. Show that

$$M^2 = \begin{bmatrix} A^2 & D \\ 0 & D^2 \end{bmatrix}$$

for some matrix D.

6. Interpret the calculation of Example 2.2.6 as a block multiplication.

7. Express the rank 1 matrix $\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 2 & 4 & 2 \end{bmatrix}$ as an outer product of two vectors.

8. Express the following in the elementary matrix notation:

(a) $\begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

9. Compute the reduced row echelon form of the following matrices and express each form as a product of elementary matrices and the original matrix.

(a) $\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 2 & 2 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & -2 \end{bmatrix}$ (d) $\begin{bmatrix} 2 & 1 \\ 0 & 1 \\ 0 & 2 \end{bmatrix}$

10. Compute the transpose and Hermitian transpose of the following matrices and determine which, if any, are symmetric or Hermitian symmetric.

(a) $\begin{bmatrix} 1 & -3 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 1 & -4 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 1 & 3 \\ 1 & 0 & 0 \\ 3 & 0 & 2 \end{bmatrix}$

11. Verify that the elementary matrix transpose law holds for 3×3 elementary matrix $E_{23}(4)$.

12. Answer True/False and give reasons:

(a) For matrix A and scalar c, $(cA)^H = cA^H$:

(b) Every diagonal matrix is symmetric.

(c) The rank of the matrix A may differ from the rank of A^T.

(d) Every diagonal matrix is Hermitian.

(e) Every tridiagonal matrix is symmetric.

13. Show that a triangular and symmetric matrix must be a diagonal matrix.

14. Show that if C has block form $C = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$, then $\operatorname{rank} C = \operatorname{rank} A + \operatorname{rank} B$.

15. Prove from definition that $(A^T)^T = A$.

16. Express the quadratic form $Q(x, y, z) = 2x^2 + y^2 + z^2 + 2xy + 4yz - 6xz$ in the matrix form $\mathbf{x}^T A\mathbf{x}$ as in Example 2.4.16.

17. Let $A = \begin{bmatrix} -2 & 1-2i \\ 0 & 3 \end{bmatrix}$ and verify that both $A^H A$ and AA^H are Hermitian.

18. Let A be an $m \times n$ matrix. Show that both $A^H A$ and AA^H are Hermitian.

19. Use Corollary 2.4.19 to prove that the outer product of any two vectors is at most a rank 1 matrix.

20. Let A be a square real matrix. Show the following:

(a) The matrix $B = \frac{1}{2}(A + A^T)$ is symmetric.

(b) The matrix $C = \frac{1}{2}(A - A^T)$ is skew-symmetric (a matrix C is *skew-symmetric* if $C^T = -C$.)

(c) The matrix A can be expressed as the sum of a symmetric matrix and a skew-symmetric matrix.

(d) With B and C as in parts (a) and (b), show that for any vector $\mathbf{x}$ of conformable size, $\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T B\mathbf{x}$.

21. Use Exercise 20 to express $A = \begin{bmatrix} 2 & 2 & -6 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$ as a sum of a symmetric and a skew-symmetric matrix. What does part (d) of this exercise say about the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$?

22. Find all 2×2 idempotent upper triangular matrices A (idempotent means $A^2 = A$).

23. Show that an $n \times n$ strictly upper triangular matrix N is nilpotent. (It might help to see what happens in a 2×2 and 3×3 case first.)

24. Let D be a diagonal matrix with distinct entries on the diagonal and $B = [b_{ij}]$ any other matrix matrix of the same size. Show that $DB = BD$ if and only if B is diagonal. Hint: Compare (i, j)th

2.5. Matrix Inverses

Definitions

We have seen that if we could make sense of "$1/A$", then we could write the solution to the linear system $A\mathbf{x} = \mathbf{b}$ as simply $\mathbf{x} = (1/A)\mathbf{b}$. We are going to tackle this problem now. First, we need a definition of the object that we are trying to uncover. Notice that "inverses" could only work on one side. For example,

$$\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = [1] = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

which suggests that both $\begin{bmatrix} 1 & 2 \end{bmatrix}$ and $\begin{bmatrix} 2 & 3 \end{bmatrix}$ should qualify as left inverses of the matrix $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$, since multiplication on the left by them results in a 1×1 identity matrix. As a matter of fact right and left inverses are studied and do have applications. But they have some unusual properties such as non-uniqueness. We are going to focus on a type of inverse that is closer to the familiar inverses in fields of numbers, namely, *two-sided* inverses. These only make sense for square matrices, so the non-square example above is ruled out.

DEFINITION 2.5.1. Let A be a square matrix. Then a *(two-sided) inverse* for A is a square matrix B of the same size as A such that $AB = I = BA$. If such a B exists, then the matrix A is said to be *invertible*.

Of course, any non-square matrix is non-invertible. Square matrices are classified as either " *singular*", i.e., non-invertible, or " *nonsingular*", i.e., invertible. Since we will mostly be concerned with two-sided inverses, the unqualified term "inverse" will be understood to mean a "two-sided inverse." Notice that this definition is actually symmetric in A and B. In other words, if B is an inverse for A, then A is an inverse for B.

Examples of Inverses

EXAMPLE 2.5.2. Show that $B = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ is an inverse for the matrix $A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$.

SOLUTION. All we have to do is check the definition. But remember that there are *two* multiplications to confirm. (We'll show later that this isn't necessary, but right now we are working strictly from the definition.) We have

$$\begin{aligned} AB &= \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 \cdot 1 - 1 \cdot 1 & 2 \cdot 1 - 1 \cdot 2 \\ -1 \cdot 1 + 1 \cdot 1 & -1 \cdot 1 + 1 \cdot 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I \end{aligned}$$

and similarly

$$BA = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 2 + 1 \cdot (-1) & 1 \cdot (-1) + 1 \cdot 1 \\ 1 \cdot 2 + 2 \cdot (-1) & 1 \cdot (-1) + 2 \cdot 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

Therefore the definition for inverse is satisfied, so that A and B work as inverses to each other. □

EXAMPLE 2.5.3. Show that the matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ cannot have an inverse.

SOLUTION. How do we get our hands on a "non-inverse"? We try an indirect approach. If A had an inverse B, then we could always find a solution to the linear system $A\mathbf{x} = \mathbf{b}$ by multiplying each side on the left by B to obtain that $BA\mathbf{x} = I\mathbf{x} = \mathbf{x} = B\mathbf{b}$, *no matter what right hand side vector* $\mathbf{b}$ we used. Yet it is easy to come up with right hand side vectors for which the system has no solution. For example, try $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Since the resulting system is clearly inconsistent, there cannot be an inverse matrix B, which is what we wanted to show. □

The moral of this last example is that it is not enough for every entry of a matrix to be nonzero for the matrix itself to be invertible. Our next example yields a gold mine of invertible matrices, namely any elementary matrix we construct.

EXAMPLE 2.5.4. Find formulas for inverses of all the elementary matrices.

SOLUTION. Remember from Corollary 2.4.4 that left multiplication by an elementary matrix is the same as performing the corresponding elementary row operation. Furthermore, from the discussion following Theorem 1.4.5 we see the following

Elementary Matrix Inverses

- E_{ij} : The elementary operation of switching the ith and jth rows is undone by applying E_{ij}. Hence
$$E_{ij}E_{ij} = E_{ij}E_{ij}I = I$$
so that E_{ij} works as its own inverse. (This is rather like -1, since $(-1)\cdot(-1) = 1$.)
- $E_i(c)$: The elementary operation of multiplying the ith row by the nonzero constant c, is undone by applying $E_i(1/c)$. Hence
$$E_i(1/c)E_i(c) = E_i(1/c)E_i(c)I = I, \text{ and}$$
$$E_i(c)E_i(1/c) = E_i(c)E_i(1/c)I = I$$
- $E_{ij}(d)$: The elementary operation of adding d times the jth row to the ith row is undone by applying $E_{ij}(-d)$. Hence
$$E_{ij}(-d)E_{ij}(d) = E_{ij}(-d)E_{ij}(d)I = I, \text{ and}$$
$$E_{ij}(d)E_{ij}(-d) = E_{ij}(-d)E_{ij}(d)I = I$$

□

Specifically, in the case of 2×2 matrices, this means, e.g., that $E_{12} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has an inverse $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, while the matrix $E_{21}(-3) = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$ has an inverse $\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} = E_{21}(3)$.

Laws of Inverses

Here are some of the basic laws of inverse calculations.

Laws of Matrix Inverses. Let A, B, C be matrices of the appropriate sizes so that the following multiplications make sense, I a suitably sized identity matrix, and c a nonzero scalar. Then

1. (Uniqueness) The matrix A has at most one inverse, henceforth denoted as A^{-1}, provided A is invertible.
2. (Double Inverse) If A is invertible, then $\left(A^{-1}\right)^{-1} = A$.
3. ($2/3$ Rule) If any two of the three matrices A, B and AB are invertible, then so is the third, and moreover $(AB)^{-1} = B^{-1}A^{-1}$.
4. If A is invertible, then $(cA)^{-1} = (1/c)A^{-1}$.
5. (Inverse/Transpose) If A is invertible, then $(A^T)^{-1} = (A^{-1})^T$ and $(A^H)^{-1} = (A^{-1})^H$.
6. (Cancellation) Suppose A is invertible. If $AB = AC$ or $BA = CA$, then $B = C$.

Notes: Observe that the $2/3$ Rule reverses order when taking the inverse of a product. This should remind you of the operation of transposing a product. A common mistake is to forget to reverse the order. Secondly, notice that the cancellation law restores something that appeared to be lost when we first discussed matrices. Yes, we can cancel a common factor from both sides of an equation, but (1) the factor must be on the same side and (2) the factor must be an invertible matrix.

Verification of Laws: Suppose that both B and C work as inverses to the matrix A. We will show that these matrices must be identical. For associativity of matrices and identity laws give that

$$B = BI = B(AC) = (BA)C = IC = C$$

Henceforth, we shall write A^{-1} for the unique (two-sided) inverse of the square matrix A, provided of course that there is an inverse at all (remember that existence of inverses is not a sure thing).

The double inverse law is a matter of examining the definition of inverse:

$$AA^{-1} = I = A^{-1}A$$

shows that A is an inverse matrix for A^{-1}. Hence, $(A^{-1})^{-1} = A$.

Now suppose that A and B are both invertible and of the same size. Use the laws of matrix arithmetic and we see that

$$AB(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

and that

$$(B^{-1}A^{-1})AB = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I$$

In other words, the matrix $B^{-1}A^{-1}$ works as an inverse for the matrix AB, which is what we wanted to show. We leave the remaining cases of the $2/3$ Rule as an exercise.

Suppose that c is nonzero and perform the calculation

$$(cA)(1/c)A^{-1} = (c/c)AA^{-1} = 1 \cdot I = I$$

A similar calculation on the other side shows that $(cA)^{-1} = (1/c)A^{-1}$.

Next, apply the transpose operator to the definition of inverse (Equation 2.5.1) and use the law of transpose products to obtain that

$$(A^{-1})^T A^T = I^T = I = A^T (A^{-1})^T$$

This shows that the definition of inverse is satisfied for $(A^{-1})^T$ relative to A^T, that is, that $(A^T)^{-1} = (A^{-1})^T$, which is the inverse/transpose law. The same argument works with Hermitian transpose in place of transpose.

Finally, if A is invertible and $AB = AC$, then multiply both sides of this equation on the left by A^{-1} to obtain that

$$A^{-1}(AB) = (A^{-1}A)B = B = A^{-1}(AC) = (A^{-1}A)C = C$$

which is the cancellation that we want.

We can now extend the power notation to negative exponents.

NOTATION 2.5.5. Let A be an invertible matrix and k a positive integer. Then we write

$$A^{-k} = A^{-1}A^{-1} \cdot \ldots \cdot A^{-1}$$

where the product is taken over k terms.

The laws of exponents that we saw earlier can now be expressed for arbitrary integers, *provided* that A is invertible. Here is an example of how we can use the various laws of arithmetic and inverses to carry out an inverse calculation.

EXAMPLE 2.5.6. Let

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

Show that $(I - A)^3 = 0$ and use this to find A^{-1}.

SOLUTION. First we check that

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

so that

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

Now we do some symbolic algebra, using the laws of matrix arithmetic:

$$0 = (I-A)^3 = (I-A)(I^2 - 2A + A^2) = I - 3A + 3A^2 - A^3$$

Subtract all terms involving A from both sides to obtain that

$$3A - 3A^2 + A^3 = A \cdot 3I - 3A^2 + A^3 = A(3I - 3A + A^2) = I$$

Since $A(3I - 3A + A^2) = (3I - 3A + A^2)A$, we see from definition of inverse that

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

□

Notice, by the way, that in the preceding example we were careful not to leave a "3" behind when we factored out A from $3A$. The reason is that $3 + 3A + A^2$ makes no sense as a sum, since one term is a scalar and the other two are matrices.

Rank and Inverse Calculation

Although we can calculate a few examples of inverses such as the last example, we really need a general procedure. So let's get right to the heart of the matter. How can we find the inverse of a matrix, or decide that none exists? Actually, we already have done all the hard work necessary to understand computing inverses. The secret is in the notion of reduced row echelon form and the attendant idea of rank. (Remember, we use elementary row operations to reduce a matrix to its reduced row echelon form. Once we have done so, the rank of the matrix is simply the number of nonzero rows in the reduced row echelon form.) Let's recall the results of Example 2.3.12:

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix} = E_{12}(-1)E_2(-1/3)E_{21}(-2)E_1(1/4)E_{12} \begin{bmatrix} 2 & -1 & 1 \\ 4 & 4 & 20 \end{bmatrix}$$

Now remove the last column from each of the matrices at the right of each side and we have this result:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = E_{12}(-1)E_2(-1/3)E_{21}(-2)E_1(1/4)E_{12} \begin{bmatrix} 2 & -1 \\ 4 & 4 \end{bmatrix}$$

This suggests that if $A = \begin{bmatrix} 2 & -1 \\ 4 & 4 \end{bmatrix}$, then

$$A^{-1} = E_{12}(-1)E_2(-1/3)E_{21}(-2)E_1(1/4)E_{12}$$

To prove this, we argue in the general case as follows: let A be an $n \times n$ matrix and suppose that by a succession of elementary row operations $E_1, E_2, \ldots, E_k$ we reduce

A to its reduced row echelon form R, which happens to be I. In the language of matrix multiplication, what we have obtained is

$$I = E_k E_{k-1} \cdot \ldots \cdot E_1 A.$$

Now let $B = E_k E_{k-1} \cdot \ldots \cdot E_1$. By repeated application of the 2/3 theorem, we see that a product of any number of invertible matrices is invertible. Since each elementary matrix is invertible, it follows that B is. Multiply both sides of the equation $I = BA$ by B^{-1} to obtain that $B^{-1}I = B^{-1} = B^{-1}BA = A$. Therefore, A is the inverse of the matrix B, hence is itself invertible.

Here's a practical trick for computing this product of elementary matrices on the fly: form what we term the *super-augmented matrix* $[A \mid I]$. Now, if we perform the elementary operation E on the super-augmented matrix we have the same result as

Super-Augmented Matrix

$$E[A \mid I] = [EA \mid EI] = [EA \mid E]$$

So the matrix occupied by the I part of the super-augmented matrix is just the product of the elementary matrices that we have used so far. Now continue applying elementary row operations until the part of the matrix originally occupied by A is reduced to the reduced row echelon form of A. We end up with this schematic picture of our calculations:

$$\left[\begin{array}{ccc} A & \mid & I \end{array}\right] \xrightarrow[E_1, E_2, \ldots, E_k]{} \left[\begin{array}{ccc} R & \mid & B \end{array}\right]$$

where R is the reduced row echelon form of A and $B = E_k E_{k-1} \cdot \ldots \cdot E_1$ is the product of the various elementary matrices we used, composed in the correct order of usage. We can summarize this discussion with the following

Inverses Algorithm : Given an $n \times n$ matrix A, to compute A^{-1}:

1. Form the super-augmented matrix $\widetilde{A} = [A \mid I_n]$.
2. Reduce the first n columns of $\widetilde{A}$ to reduced row echelon form by performing elementary operations on the matrix $\widetilde{A}$ resulting in the matrix $[R \mid B]$.
3. If $R = I_n$ then set $A^{-1} = B$, else A is singular and A^{-1} does not exist.

EXAMPLE 2.5.7. Use the Inverses Algorithm to compute the inverse of Example 2.2.5,

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

SOLUTION. Notice that this matrix is already upper triangular. Therefore, as in Gaussian elimination, it is a bit more efficient to start with the bottom pivot and clear out entries above in reverse order. So we compute

$$[A \mid I_3] = \begin{bmatrix} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow[E_{23}(-1)]{} \begin{bmatrix} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \xrightarrow[E_{1,2}(-2)]{} \begin{bmatrix} 1 & 0 & 0 & 1 & -2 & 2 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

So we conclude that A is indeed invertible and

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

□

There is a simple formula for the inverse of a general 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Set $D = ad - bc$. It is easy to verify that if $D \neq 0$, then

$$A^{-1} = \frac{1}{D} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

EXAMPLE 2.5.8. Use the two by two inverse formula to find the inverse of the matrix $A = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$, and verify that the same answer results if we use the inverses algorithm.

SOLUTION. First we apply the inverses algorithm.

$$\begin{bmatrix} 1 & -1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix} \xrightarrow[E_{21}(-1)]{} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 3 & -1 & 1 \end{bmatrix} \xrightarrow[E3(1/3)]{} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1/3 & 1/3 \end{bmatrix}$$
$$\xrightarrow[E_{12}(1)]{} \begin{bmatrix} 1 & 0 & 2/3 & 1/3 \\ 0 & 1 & -1/3 & 1/3 \end{bmatrix}$$

Thus we have found that $\begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}^{-1} = \frac{1}{3} \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}$.

To apply the inverse formula, calculate $D = 1 \cdot 2 - 1 \cdot (-1) = 3$. Swap diagonal entries of A, negate the off-diagonal entries and divide by D to get the same result as we have just obtained in the preceding equation for the inverse. □

The formula of the preceding example is well worth memorizing, since we will frequently need to find the inverse of a 2×2 matrix. Notice that in order for it to make sense, we have to have D nonzero. The number D is called the *determinant* of the matrix A. We will have more to say about this number in the next section. In our current example it is fairly easy to see why A must have $D \neq 0$ in order for its inverse to exist if we look ahead to the next theorem. Notice in the above elementary operation calculations that if $D = 0$ then elementary operations on A lead to a matrix with a row of zeros. Therefore, the rank of A will be smaller than 2. Here is a summary of our current knowledge of the invertibility of a square matrix.

Conditions for Invertibility

THEOREM 2.5.9. *The following are equivalent conditions on the square $n \times n$ matrix A:*

1. *The matrix A is invertible.*
2. *There is a square matrix B such that $BA = I$.*
3. *The linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every right hand side vector $\mathbf{b}$.*
4. *The linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution for some right hand side vector $\mathbf{b}$.*
5. *The linear system $A\mathbf{x} = 0$ has only the trivial solution.*

6. $\operatorname{rank} A = n$.
7. *The reduced row echelon form of* A *is* I_n.
8. *The matrix* A *is a product of elementary matrices.*

PROOF. The method of proof here is to show that each condition implies the next, and that the last condition implies the first. This connects all the conditions in a circle, so that any one condition will imply any other and therefore all are equivalent to each other. Here is our chain of reasoning:

(1) implies (2): Assume A is invertible. Then the choice $B = A^{-1}$ certainly satisfies condition (2).

(2) implies (3): Assume (2) is true. Given a system $A\mathbf{x} = \mathbf{b}$, we can multiply both sides on the left by B to get that $\mathbf{x} = I\mathbf{x} = BA\mathbf{x} = B\mathbf{b}$. So there is only one solution, if any. On the other hand, if the system were inconsistent then we would have $\operatorname{rank} A < n$. By Corollary 2.4.19 $\operatorname{rank} BA < n$, contradicting the fact that $\operatorname{rank} I_n = n$. Hence, there is a solution, which proves (3).

(3) implies (4): This statement is obvious.

(4) implies (5): Assume (4) is true. Say the unique solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x}_0$. If the system $A\mathbf{x} = 0$ had a nontrivial solution, say $\mathbf{z}$, then we could add $\mathbf{z}$ to $\mathbf{x}_0$ to obtain a different solution $\mathbf{x}_0 + \mathbf{z}$ of the system $A\mathbf{x} = \mathbf{b}$ (check: $A(\mathbf{z} + \mathbf{x}_0) = A\mathbf{z} + A\mathbf{x}_0 = 0 + \mathbf{b} = \mathbf{b}$.) This is impossible since (4) is true, so (5) follows.

(5) implies (6): Assume (5) is true. We know from Theorem 1.4.15 of Chapter 1 that the consistent system $A\mathbf{x} = 0$ has a unique solution precisely when the rank of A is n. Hence (6) must be true.

(6) implies (7): Assume (6) is true. The reduced row echelon form of A is the same size as A, that is $n \times n$, and must have a row pivot entry 1 in every row which must be the only nonzero entry in its column. This exactly describes the matrix I_n, so that (7) is true.

(7) implies (8): Assume (7) is true. We know that the matrix A is reduced to its reduced row echelon form by applying a sequence of elementary operations, or what amounts to the same thing, by multiplying the matrix A on the left by elementary matrices $E_1, E_2, \ldots, E_k$, say. Then $E_1 E_2 \ldots E_k A = I$. But we know from Example 2.5.4 that each elementary matrix is invertible and that their inverses are themselves elementary matrices. By successive multiplications on the left we obtain that $A = E_k^{-1} E_{k-1}^{-1} \ldots E_1^{-1} I$, showing that A is a product of elementary matrices which is condition (8).

(8) implies (1): Assume (8) is true. Repeated application of the $2/3$ Inverses Rule shows that the product of any number of invertible matrices is itself invertible. Since elementary matrices are invertible, condition (1) must be true. □

There is an interesting consequence to Theorem 2.5.9 that has been found to be useful in some contexts. It's an either/or statement, so it will always have something to say about any square linear system. This type of statement is sometimes called a *Fredholm alternative*. Many theorems go by this name, and we'll state another one in Chapter 5. Notice that a matrix is not invertible if and only one of the conditions of the Theorem fail. Certainly it is true that either a square matrix is invertible or not invertible. That's

Fredholm Alternative

all this Fredholm alternative really says, but it uses the equivalent conditions (3) and (5) of Theorem 2.5.9 to say it in a different way:

COROLLARY 2.5.10. *Given a square linear system $A\mathbf{x} = \mathbf{b}$, either the system has a unique solution for every right hand side vector $\mathbf{b}$ or there is a nonzero solution $\mathbf{x} = \mathbf{x}_0$ to the homogeneous system $A\mathbf{x} = 0$.*

We conclude this section with an application of the matrix algebra developed so far to the problem of solving nonlinear equations. Although we focus on two equations in two unknowns, the same ideas can be extended to any number of equations in as many unknowns.

Recall that we could solve the one variable (usually nonlinear) equation $f(x) = 0$ for a solution point x_1 at which $f(x_1) = 0$ from a given "nearby" point x_0 by setting $dx = x_1 - x_0$, and assuming that the change in f is

$$\begin{aligned}\Delta f &= f(x_1) - f(x_0) = -f(x_0) \\ &\approx df = f'(x_0)\,dx = f'(x_0)(x_1 - x_0).\end{aligned}$$

Now solve for x_1 in the equation $-f(x_0) = f'(x_0)(x_1 - x_0)$ and get the equation

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Replace 1 by $n+1$ and 0 by n to obtain the famous Newton formula:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{2.5.1}$$

The idea is to start with x_0, use the formula to get x_1 and if $f(x_1)$ is not close enough to 0, then repeat the calculation with x_1 in place of x_0, and so forth until a satisfactory value of $x = x_n$ is reached. How does this relate to a two variable problem? We illustrate the basic idea in two variables.

Newton's Method for Systems

EXAMPLE 2.5.11. Describe concisely an algorithm analogous to Newton's method in one variable to solve the two variable problem

$$\begin{aligned} x^2 + y^2 + \sin(xy) &= 1 \\ xe^{x+y} - y\sin(x+y) &= 0 \end{aligned}$$

SOLUTION. Our problem can be written as a system of two (nonlinear) equations in two unknowns, namely

$$\begin{aligned} f(x,y) &= x^2 + y^2 + \sin(xy) - 1 = 0 \\ g(x,y) &= xe^{x+y} - y\sin(x+y) = 0 \end{aligned}$$

Now we can pull the same trick with differentials as in the one variable problem by setting $dx = x_1 - x_0, dy = y_1 - y_0$, where $f(x_1, y_1) = 0$, approximating the change in both f and g by differentials, and recalling the definition of these differentials in terms of partial derivatives. This leads to a system

$$\begin{aligned} f_x(x_0,y_0)\,dx + f_y(x_0,y_0)\,dy &= -f((x_0,y_0) \\ g_x(x_0,y_0)\,dx + g_y(x_0,y_0)\,dy &= -g((x_0,y_0) \end{aligned}$$

Next, write everything in vector style, say

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} f(\mathbf{x}) \\ g(\mathbf{x}) \end{bmatrix}, \quad \mathbf{x}^{(0)} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}, \quad \mathbf{x}^{(1)} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

Now we can write the *vector* differentials in the forms

$$d\mathbf{F} = \begin{bmatrix} df \\ dg \end{bmatrix}, \quad \text{and } d\mathbf{x} = \begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} x_1 - x_0 \\ y_1 - x_0 \end{bmatrix} = \mathbf{x}^{(1)} - \mathbf{x}^{(0)}$$

The original Newton equations now look like a matrix multiplication involving $d\,\mathbf{x}, \mathbf{F}$ and a matrix of derivatives of $\mathbf{F}$, namely the so-called Jacobian matrix

$$J_{\mathbf{F}}(x_0, y_0) = \begin{bmatrix} f_x(x_0, y_0) & f_y(x_0, y_0) \\ g_x(x_0, y_0) & g_y(x_0, y_0) \end{bmatrix}$$

Specifically, we see from the definition of matrix multiplication that the Newton equations are equivalent to the vector equations

$$d\mathbf{F} = J_{\mathbf{F}}(\mathbf{x}_0)\, d\,\mathbf{x} = -\mathbf{F}(\mathbf{x}^{(0)})$$

Thus we obtain that if the Jacobian matrix is invertible then

$$\mathbf{x}^{(1)} - \mathbf{x}^{(0)} = d\,\mathbf{x} = -J_{\mathbf{F}}(\mathbf{x}^{(0)})^{-1}\,\mathbf{F}(\mathbf{x}^{(0)})$$

whence by adding $\mathbf{x}_0$ to both sides we see that

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - J_{\mathbf{F}}(\mathbf{x}^{(0)})^{-1}\,\mathbf{F}(\mathbf{x}^{(0)})$$

Now replace 1 by $n+1$ and 0 by n to obtain the ever famous Newton formula in vector form:

Newton's Formula in Vector Form

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - J_{\mathbf{F}}(\mathbf{x}^{(n)})^{-1}\,\mathbf{F}(\mathbf{x}^{(n)})$$

This is a beautiful analogy to the Newton formula of (2.5.1) which would not have been possible without the language of vectors and matrices. □

2.5 Exercises

1. Find the inverse of the following matrices, or show that it does not exist:

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

(d) $\begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 2+i \\ i & 2 \end{bmatrix}$ (f) $\begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}$

2. Given the matrix A and vector $\mathbf{b}$, find the inverse of the matrix A and use this to solve the system $A\mathbf{x} = \mathbf{b}$, where

(a)$A = \begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$ (b)$A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$

3. Solve the following systems by inverting the coefficient matrix of the system.

(a) $\begin{aligned} 2x+3y &= 7 \\ x+2y &= -2 \end{aligned}$ (b) $\begin{aligned} 3x_1+6x_2-x_3 &= -4 \\ -2x_1+x_2+x_3 &= 3 \\ x_3 &= 1 \end{aligned}$ (c) $\begin{aligned} x_1+x_2 &= -2 \\ 5x_1+2x_2 &= 5 \end{aligned}$

4. Find 2×2 matrices A, B and C such that $AB = CA$ but $B \neq C$.

5. Find $A^{-1}B$ if $A = \begin{bmatrix} 1 & 2 & -3 \\ 0 & -1 & 1 \\ 2 & 5 & -6 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & -1 & 1 & 1 \\ 2 & 0 & -6 & 0 \end{bmatrix}$.

6. Determine the inverses for the following matrices in terms of the parameter c and conditions on c for which the matrix has an inverse.

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

7. Prove from the definition that if a square matrix A satisfies the equation $A^3 - 2A + 3I = 0$, then the matrix A must be invertible.

8. Express the following matrices and their inverses in the notation of elementary matrices.

(a) $\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

9. Show directly from the definition of inverse that the two by two inverse formula gives the inverse of a 2×2 matrix.

10. Assume that the product of invertible matrices is invertible and deduce that if A and B are invertible matrices of the same size and both B and AB are invertible, then so is A.

11. Let A be an invertible matrix.

(a) Show that if the product of matrices AB is defined, then $\operatorname{rank}(AB) = \operatorname{rank}(B)$.

(b) Show that if BA is defined, then $\operatorname{rank}(BA) = \operatorname{rank}(B)$.

12. Suppose the matrix $M = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$, where the blocks A and C are invertible matrices. Find a formula for M^{-1} in terms of A, B, C. *Hint:* Assume M^{-1} has the same form as M and solve for the blocks in M using $MM^{-1} = I$.

13. Verify that for any square matrix N and positive integer k that $(I+N+N^2+\ldots+N^{k-1})(I-N) = I - N^k$.

14. Use the Exercise 13 to find a formula for the inverse of the matrix $I - N$, where N is nilpotent, i.e., $N^k = 0$ for some positive k. Test this formula on the matrices

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

15. Use a calculator to apply the Newton formula (Equation (2.5.1)) to the one variable problem

$$x^2 + \sin^2(x) = 1$$

starting with $x_0 = 1/2$ and performing three iterations. Remember to first rewrite the equation in the form $f(x) = 0$. What is the value of $f(x_3)$?

16. Solve the nonlinear system of equations of Example 2.5.11 by using four iterations of the vector Newton formula (2.5), starting with the initial guess $\mathbf{x_0} = (1/2, 1/2)$. How small is $F(\mathbf{x_4})$? You will need a reasonably competent calculator or a computer algebra program to do this exercise.

17. Find the minimum value of the function $F(x, y) = (x^2 + y + 1)^2 + x^4 + y^4$ by using the Newton method to find critical points of the function $F(x, y)$, i.e., points where the system $f(x, y) = F_x(x, y) = 0$ and $g(x, y) = F_y(x, y) = 0$.

18. Show that if the product of matrices BA is defined and A is invertible, then $\operatorname{rank}(BA) = \operatorname{rank}(B)$

19. Show that if D is an $n \times n$ diagonal matrix with nonzero diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$, then the inverse of D is the $n \times n$ diagonal matrix with diagonal entries $1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_n$.

2.6. Basic Properties of Determinants

What are they?

Many students have already had some experience with determinants and may have used them in high school to solve square systems of equations. Why have we waited until now to introduce them? In point of fact, they are not really the best tool for solving systems. That distinction goes to Gaussian elimination. Were it not for the *theoretical* usefulness of determinants they might be consigned to a footnote in introductory linear algebra texts as an historical artifact of linear algebra.

To motivate determinants, consider Example 2.5.8. Something remarkable happened in that example. Not only were we able to find a formula for the inverse of a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, but we were able to compute a single number $D = ad - bc$ that told us whether A was invertible or not. The condition of non-invertibility, namely that $D = 0$, has a very simple interpretation: this happens exactly when one row of A is a multiple of the other, since the example showed that this is when elementary operations use the first row to zero out the second row. Can we extend this idea? Is there a single number that will tell us whether or not there are dependencies among the rows of the square matrix A that cause its rank to be smaller than its row size? The answer is yes. This is exactly what determinants were invented for. The concept of determinant is subtle and not intuitive, and researchers had to accumulate a large body of experience before they were able to formulate a "correct" definition for this number. There are alternate definitions of determinants, but the following will suit our purposes. It is sometimes referred to as "expansion down the first column."

DEFINITION 2.6.1. The *determinant* of a square matrix $n \times n$ matrix $A = [a_{ij}]$ is the scalar quantity $\det A$ defined recursively as follows: if $n = 1$ then $\det A = a_{11}$; otherwise, we suppose that determinants are defined for all square matrices of size less than n and specify that

$$\begin{aligned} \det A &= \sum_{k=1}^{n} a_{k1}(-1)^{k+1} M_{k1}(A) \\ &= a_{11}M_{11}(A) - a_{21}M_{21}(A) + \ldots + (-1)^{n+1}a_{n1}M_{n1}(A) \end{aligned}$$

where $M_{ij}(A)$ is the determinant of the $(n-1) \times (n-1)$ matrix obtained from A by deleting the ith row and jth column of A.

Caution: The determinant of a matrix A is a scalar number. It is *not* a matrix quantity.

EXAMPLE 2.6.2. Describe the quantities $M_{21}(A)$ and $M_{22}(A)$ where

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

SOLUTION. If we erase the second row and first column of A we obtain something like

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

Now collapse the remaining entries together to obtain the matrix

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

Therefore

$$M_{21}(A) = \det \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}.$$

Similarly, erase the second row and column of A to obtain

$$\begin{bmatrix} 2 & & 0 \\ & & \\ 0 & & 2 \end{bmatrix}.$$

Now collapse the remaining entries together to obtain

$$M_{22}(A) = \det \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

□

Now how do we calculate these determinants? Part (b) of the next example answers the question.

EXAMPLE 2.6.3. Use the definition to compute the determinants of the following matrices:

$$\text{(a) } [-4] \qquad \text{(b) } \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \text{(c) } \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

SOLUTION. (a) From the first part of the definition we see that

$$\det[-4] = -4$$

For (b) we set $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and use the formula of the definition to obtain that

$$\begin{aligned} \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= a_{11}M_{11}(A) - a_{21}M_{21}(A) \\ &= a \det[d] - c \det[b] \\ &= ad - cb \end{aligned}$$

This calculation gives a handy formula for the determinant of a 2×2 matrix. For (c) use the definition to obtain that

$$\begin{aligned} \det \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & 1 & 2 \end{bmatrix} &= 2 \det \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} - 1 \det \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix} + 0 \det \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \\ &= 2(1 \cdot 2 - 1 \cdot (-1)) - 1(1 \cdot 2 - 1 \cdot 0) + 0(1 \cdot (-1) - 1 \cdot 0) \\ &= 2 \cdot 3 - 1 \cdot 2 + 0 \cdot (-1) \\ &= 4 \end{aligned}$$

A point worth observing here is that we didn't really have to calculate the determinant of any matrix if it is multiplied by a zero. Hence, the more zeros our matrix has, the easier we expect the determinant calculation to be! □

NOTATION 2.6.4. Another common symbol for $\det A$ is $|A|$, which is also written with respect to the elements of A by suppressing matrix brackets:

$$\det A = |A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

This notation invites a certain oddity, if not abuse, of language: we sometimes refer to things like the "second row" or "$(2,3)$th element" or the "size" of the determinant. Yet the determinant is only a number and doesn't really have rows or entries or a size. Rather, it is the underlying matrix whose determinant is being calculated that has these properties. So be careful of this notation; we plan to use it frequently because it's handy, but you should bear in mind that determinants and matrices are *not* the same thing! Another reason that this notation can be tricky is the case of a one dimensional matrix, say $A = [a_{11}]$. Here it is definitely *not* a good idea to forget the brackets, since we already understand $|a_{11}|$ to be the absolute value of the scalar a_{11}, a nonnegative number. In the 1×1 case use $|[a_{11}]|$ for the determinant, which is just the number a_{11} and may be positive or negative.

Minors and Cofactors

NOTATION 2.6.5. The number $M_{ij}(A)$ is called the (i, j)th *minor* of the matrix A. If we collect the sign term in the definition of determinant together with the minor we obtain the (i, j)th *cofactor* $A_{ij} = (-1)^{i+j} M_{ij}(A)$ of the matrix A. In the terminology of cofactors,

$$\det A = \sum_{k=1}^{n} a_{k1} A_{k1}$$

Laws of Determinants

Our primary goal here is to show that determinants have the magical property we promised: a matrix is singular exactly when its determinant is 0. Along the way we will examine some useful properties of determinants. There is a lot of clever algebra that can be done here; we will try to keep matters straightforward (if that's possible with determinants). In order to focus on the main ideas, we will place most of the proofs of key facts at the end of the next section for optional reading. Also, a concise summary of the basic determinantal laws is given at the end of this section. Unless otherwise stated, we assume throughout this section that matrices are square, and that $A = [a_{ij}]$ is an $n \times n$ matrix.

For starters, let's observe that it's very easy to calculate the determinant of upper triangular matrices. Let A be such a matrix. Then $a_{k1} = 0$ if $k > 1$, so

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix}$$
$$= \cdots = a_{11} \cdot a_{22} \cdot \ldots \cdot a_{nn}$$

Hence we have established our first determinantal law:

D1: If A is an upper triangular matrix, then the determinant of A is the product of all the diagonal elements of A.

EXAMPLE 2.6.6. Compute $D = \begin{vmatrix} 4 & 4 & 1 & 1 \\ 0 & -1 & 2 & 3 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 2 \end{vmatrix}$ and $|I_n| = \det I_n$.

SOLUTION. By D1 we can do this at a glance: $D = 4 \cdot (-1) \cdot 2 \cdot 2 = -16$. Since I_n is diagonal, it is certainly upper triangular. Moreover, the entries down the diagonal of this matrix are 1's, so D1 implies that $|I_n| = 1$. $\square$

Next, suppose that we notice a common factor of the scalar c in a row, say for convenience, the first one. How does this affect the determinantal calculation? In the case of a 1×1 determinant, we could simply factor it out of the original determinant. The general situation is covered by this law:

D2: If B is obtained from A by multiplying one row of A by the scalar c, then $\det B = c \cdot \det A$.

Here is a simple illustration:

EXAMPLE 2.6.7. Compute $D = \begin{vmatrix} 5 & 0 & 10 \\ 5 & 5 & 5 \\ 0 & 0 & 2 \end{vmatrix}$.

SOLUTION. Put another way, D2 says that scalars may be factored out of individual rows of a determinant. So use D2 on the first and second rows and then use definition of determinant to obtain

$$\begin{vmatrix} 5 & 0 & 10 \\ 5 & 5 & 5 \\ 0 & 0 & 2 \end{vmatrix} = 5 \cdot \begin{vmatrix} 1 & 0 & 2 \\ 5 & 5 & 5 \\ 0 & 0 & 2 \end{vmatrix} = 5 \cdot 5 \cdot \begin{vmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ 0 & 0 & 2 \end{vmatrix}$$
$$= 25 \cdot \left(1 \cdot \begin{vmatrix} 1 & 1 \\ 0 & 2 \end{vmatrix} - 1 \cdot \begin{vmatrix} 0 & 2 \\ 0 & 2 \end{vmatrix} + 0 \cdot \begin{vmatrix} 0 & 2 \\ 1 & 1 \end{vmatrix} \right) = 50$$

One can easily check that this is the same answer we get by working the determinant directly from definition. □

Next, suppose we interchange two rows of a determinant. Then we have the following:

D3: If B is obtained from A by interchanging two rows of A, then $\det B = -\det A$.

EXAMPLE 2.6.8. Use D3 to show the following handy fact: if a determinant has a repeated row, then it must be 0.

SOLUTION. Suppose that the ith and jth rows of the matrix A are identical, and B is obtained by switching these two rows of A. Clearly $B = A$. Yet, according to D3, $\det B = -\det A$. It follows that $\det A = -\det A$, i.e., if we add $\det A$ to both sides, $2 \cdot \det A = 0$, so that $\det A = 0$, which is what we wanted to show. □

Now we ask what happens to a determinant if we add a multiple of one row to another. The answer is as follows.

D4: If B is obtained from A by adding a multiple of one row of A to another row of A, then $\det B = \det A$.

EXAMPLE 2.6.9. Compute $D = \begin{vmatrix} 1 & 4 & 1 & 1 \\ 1 & -1 & 2 & 3 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 1 & 2 \end{vmatrix}$.

SOLUTION. What D4 really says is that any elementary row operation $E_{ij}(c)$ can be applied to the matrix behind a determinant and the determinant will be unchanged. So in this case, add -1 times the first row to the second and $-1/2$ times the third row to the fourth, then apply D1 to obtain

$$\begin{vmatrix} 1 & 4 & 1 & 1 \\ 1 & -1 & 2 & 3 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 1 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 4 & 1 & 1 \\ 0 & -5 & 1 & 2 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 1/2 \end{vmatrix} = 1 \cdot (-5) \cdot 2 \cdot \frac{1}{2} = -5$$

□

EXAMPLE 2.6.10. Use D3 to show that a matrix with a row of zeros has zero determinant.

SOLUTION. Suppose A has a row of zeros. Add any other row of the matrix A to this zero row to obtain a matrix B with repeated rows. □

We now have enough machinery to establish the most important property of determinants. First of all, we can restate laws D2-D4 in the language of elementary matrices as follows:

Determinant of Elementary Matrices

- D2: $\det(E_i(c)A) = c \cdot \det A$ (remember that for $E_i(c)$ to be an elementary matrix, $c \neq 0$.)
- D3: $\det(E_{ij}A) = -\det A$
- D4: $\det(E_{ij}(s)A) = \det A$

Apply a sequence of elementary row operations on the $n \times n$ matrix A to reduce it to its reduced row echelon form R, or equivalently, multiply A on the left by elementary matrices $E_1, E_2, \ldots, E_k$ and obtain

$$R = E_1 E_2 \ldots E_k A$$

Take determinant of both sides to obtain

$$\det R = \det(E_1 E_2 \ldots E_k A) = \pm(\text{nonzeroconstant}) \cdot \det A$$

Therefore, $\det A = 0$ precisely when $\det R = 0$. Now the reduced row echelon form of A is certainly upper triangular. In fact, it is guaranteed to have zeros on the diagonal, and therefore have zero determinant by D1, unless $\operatorname{rank} A = n$, in which case $R = I_n$. According to Theorem 2.5.9 this happens precisely when A is invertible. Thus we have shown:

D5: The matrix A is invertible if and only if $\det A \neq 0$.

EXAMPLE 2.6.11. Determine if the following matrices are invertible or not without actually finding the inverse:

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

SOLUTION. Compute determinants:

$$\begin{vmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & 1 & 2 \end{vmatrix} = 2 \begin{vmatrix} 1 & -1 \\ 1 & 2 \end{vmatrix} - 1 \begin{vmatrix} 1 & 0 \\ 1 & 2 \end{vmatrix}$$
$$= 2 \cdot 3 - 2 = 4$$

and

$$\begin{vmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & -1 & 2 \end{vmatrix} = 2 \begin{vmatrix} 1 & -1 \\ -1 & 2 \end{vmatrix} - 1 \begin{vmatrix} 1 & 0 \\ -1 & 2 \end{vmatrix}$$
$$= 2 \cdot 1 - 1 \cdot 2 = 0$$

Hence by D5 matrix (a) is invertible and matrix (b) is not invertible. □

There are two more surprising properties of determinants that we now discuss. Their proofs involve using determinantal properties of elementary matrices (see the next section for details).

D6: Given matrices A, B of the same size,

$$\det AB = \det A \det B$$

EXAMPLE 2.6.12. Verify D6 in the case that $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$. How do $\det(A+B)$ and $\det A + \det B$ compare in this case?

SOLUTION. We have easily that $\det A = 1$ and $\det B = 2$. Therefore, $\det A + \det B = 1 + 2 = 3$, while $\det A \cdot \det B = 1 \cdot 2 = 2$. On the other hand

$$AB = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}$$
$$A + B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$

so that $\det AB = 2 \cdot 3 - 4 \cdot 1 = 2 = \det A \cdot \det B$, as expected. On the other hand we have that $\det(A+B) = 3 \cdot 2 - 1 \cdot 1 = 5 \neq \det A + \det B$. □

This example raises a very important point.

Caution: In general, $\det A + \det B \neq \det(A+B)$, though there are occasional exceptions.

In other words, determinants do not distribute over sums. (It is true, however, that the determinant is additive in *one row at a time*. See the proof of D4 for details.)

Finally, we ask how $\det A^T$ compares to $\det A$. Try a simple case like the 2×2 and we discover that there seems to be no difference in determinant. This is exactly what happens in general.

D7: For all square matrices A, $\det A^T = \det A$.

EXAMPLE 2.6.13. Compute $D = \begin{vmatrix} 4 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 \\ 1 & 2 & -2 & 0 \\ 1 & 0 & 1 & 2 \end{vmatrix}$.

SOLUTION. By D7 and D1 we can do this at a glance: $D = 4 \cdot 1 \cdot (-2) \cdot 2 = -16$. □

D7 is a very useful fact. Let's look at it from this point of view: transposing a matrix interchanges the rows and columns of the matrix. Therefore, everything that we have said about rows of determinants applies equally well to the columns, *including the definition of determinant itself!* Therefore, we could have given the definition of determinant in terms of expanding across the first row instead of down the first column and gotten the same answers. Likewise, we could perform elementary column operations instead of row operations and get the same results as D2-D4. Furthermore, the determinant of a lower triangular matrix is the product of its diagonal elements thanks to D7+D1. By interchanging rows or columns then expanding by first row or column, we see that the same effect is obtained by simply expanding the determinant down any column or across any row. We have to alternate signs starting with the sign $(-1)^{i+j}$ of the first term we use.

Now we can really put it all together and compute determinants to our heart's content with a good deal less effort than the original definition specified. We can use D1-D4 in particular to make a determinant calculation no worse than Gaussian elimination in the amount of work we have to do. We simply reduce a matrix to triangular form by elementary operations, then take the product of the diagonal terms.

EXAMPLE 2.6.14. Calculate $D = \begin{vmatrix} 3 & 0 & 6 & 6 \\ 1 & 0 & 2 & 1 \\ 2 & 0 & 0 & 1 \\ -1 & 2 & 0 & 0 \end{vmatrix}$.

SOLUTION. We are going to do this calculation two ways. We may as well use the same elementary operation notation that we have employed in Gaussian elimination. The only difference is that we have equality instead of arrows, provided that we modify the value of the new determinant in accordance with the laws D1-D3. So here is the straightforward method:

$$D = 3\begin{vmatrix} 1 & 0 & 2 & 2 \\ 1 & 0 & 2 & 1 \\ 2 & 0 & 0 & 1 \\ -1 & 2 & 0 & 0 \end{vmatrix} \underset{\substack{E_{21}(-1) \\ E_{31}(-2) \\ E_{41}(1)}}{=} 3\begin{vmatrix} 1 & 0 & 2 & 2 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -4 & -3 \\ 0 & 2 & 2 & 2 \end{vmatrix}$$

$$\underset{E_{24}}{=} -3\begin{vmatrix} 1 & 0 & 2 & 2 \\ 0 & 2 & 2 & 2 \\ 0 & 0 & -4 & -3 \\ 0 & 0 & 0 & -1 \end{vmatrix} = -24$$

Here is another approach: let's expand the determinant down the second column, since it is mostly 0's. Remember that the sign in front of the first minor must be $(-1)^{1+2} = -1$. Also, the coefficients of the first three minors are 0, so we need only write down the last one in the second column:

$$D = +2\begin{vmatrix} 3 & 6 & 6 \\ 1 & 2 & 1 \\ 2 & 0 & 1 \end{vmatrix}$$

Expand down the second column again:

$$\begin{aligned} D &= 2\left(-6\begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} + 2\begin{vmatrix} 3 & 6 \\ 2 & 1 \end{vmatrix}\right). \\ &= 2(-6 \cdot (-1) + 2 \cdot (-9)) = -24 \end{aligned}$$

□

Summary of Determinantal Laws

Now that our list of the basic laws of determinants is complete, we record them in a concise format which includes two laws (**D7** and **D8**) to be discussed in the next section.

Laws of Determinants. Let A, B be $n \times n$ matrices. Then

- **D1:** If A is an upper triangular matrix, then $\det A$ is the product of all the diagonal elements of A.
- **D2:** $\det(E_i(c)A) = c \cdot \det A$ (here $c \neq 0$.)
- **D3:** $\det(E_{ij}A) = -\det A$.
- **D4:** $\det(E_{ij}(s)A) = \det A$.
- **D5:** The matrix A is invertible if and only if $\det A \neq 0$.
- **D6:** $\det AB = \det A \det B$.
- **D7:** $\det A^T = \det A$.
- **D8:** $A \operatorname{adj} A = (\operatorname{adj} A)A = (\det A)I$.
- **D9:** If $\det A \neq 0$, then $A^{-1} = \frac{1}{\det A} \operatorname{adj} A$.

2.6 Exercises

1. Compute all minors and cofactors for these matrices:

$$\text{(a)} \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & 1 \\ 3 & 0 & 1 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \qquad \text{(c)} \begin{bmatrix} 1 & 1-i \\ i & 0 \end{bmatrix}$$

2. Compute these determinants and determine which of these matrices whose are invertible.

$$\text{(a)} \begin{vmatrix} 2 & -1 \\ 1 & 1 \end{vmatrix} \quad \text{(b)} \begin{vmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1+i \end{vmatrix} \quad \text{(c)} \begin{vmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 2 & 0 & 1 \end{vmatrix} \quad \text{(d)} \begin{vmatrix} 1 & -1 & 4 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 2 & 7 \\ -2 & 3 & 4 & 6 \end{vmatrix}$$

$$\text{(e)} \begin{vmatrix} 1 & -1 & 4 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 2 & 7 \\ -2 & 3 & 4 & 6 \end{vmatrix} \quad \text{(f)} \begin{vmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 \\ 0 & 0 & 1 & 3 \\ 1 & 1 & 2 & 1 \end{vmatrix} \quad \text{(g)} \begin{vmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 2 & 0 \end{vmatrix} \quad \text{(h)} \begin{vmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{vmatrix}$$

3. Verify the determinants laws D6 and D7 for the following matrices:

$$A = \begin{bmatrix} -2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

4. Verify that $\begin{vmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & e & f \\ 0 & 0 & g & h \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} e & f \\ g & h \end{vmatrix}$.

5. Use determinants to find conditions on the parameters in these matrices under which the matrices are invertible.

$$\text{(a)} \begin{bmatrix} a & 1 \\ ab & 1 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 1 & 1 & -1 \\ 1 & c & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} \lambda-1 & 0 & 0 \\ 1 & \lambda-2 & 1 \\ 3 & 1 & \lambda-1 \end{bmatrix}$$

(d) $\lambda I_2 - \begin{bmatrix} 0 & 1 \\ -c_0 & -c_1 \end{bmatrix}$

6. Let

$$V = \begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{bmatrix}$$

(Such a matrix is called a *Vandermonde* matrix.)) Express $\det V$ as a product of factors $(x_j - x_k)$. ***Hint:*** Use elementary operations to clear the first column and factor out as many $(x_j - x_k)$ as possible in the resulting determinant.

7. Use the determinantal law D6 to show that $\det A \det A^{-1} = 1$ if A is invertible.

8. Show by example that $\det A^H \neq \det A$ and prove that in general $\det A^H = \overline{\det A}$.

9. Use the determinantal laws to show that any matrix with a row of zeros has zero determinant.

10. If A is a 5×5 matrix, then in terms of $\det(A)$, what can we say about $\det(-2A)$? $\det(A^{-1})$? Explain.

11. Show that if

$$M = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

then $\det M = \det A \cdot \det C$. ***Hint:*** Use row operations to make the diagonal submatrices triangular.

12. Prove that if A is $n \times n$, then $\det(-A) = (-1)^n \det A$.

13. Let A be a skew-symmetric matrix, that is, $A^T = -A$. Show that A must be singular.

14. Let J_n be the $n \times n$ counteridentity , that is, J_n is a square matrix with ones along the counterdiagonal (the diagonal that starts in the lower left corner and ends in the upper right corner), and zeros elsewhere.

(a) Prove that $J_n^2 = I_n$.

(b) Prove that $J_n^T = J_n$.

(c) Find a formula for $\det J_n$.

15. Show that the *companion matrix* of the polynomial $f(x) = c_0 + c_1 x + \cdots c_{n-1} x^{n-1} + x^n$, that is,

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

is invertible if and only if $c_0 \neq 0$.

2.7. *Applications and Proofs for Determinants

There are two fundamental applications of determinants that we develop in this section. The first is the derivation of an explicit formula for the inverse of a matrix in terms of its coefficients and determinant, which extends the 2×2 Example 2.5.8 to matrices of all sizes. From this example, one might wonder if a similar calculation could be done for any matrix. We will see that it can. The second application is something that many have already seen in high school algebra, at least for 2×2 and 3×3 systems: Cramer's Rule gives a way of solving square systems, provided that they have a unique solution.

An Inverse Formula

Let $A = [a_{ij}]$ be $n \times n$. We have already seen that we can expand the determinant of A down any column of A (see the discussion following Example 2.6.13). These lead to cofactor formulas for each column number j:

$$\det A = \sum_{k=1}^{n} a_{kj} A_{kj} = \sum_{k=1}^{n} A_{kj} a_{kj}$$

This formula resembles a matrix multiplication formula. Consider the slightly altered sum

$$\sum_{k=1}^{n} A_{ki} a_{kj} = A_{1i} a_{1j} + A_{2i} a_{2j} + \ldots + A_{ni} a_{nj}$$

The key to understanding this expression is to realize that it is exactly what we would get if we replaced the ith column of the matrix A by its jth column and then computed the determinant of the resulting matrix by expansion down the ith column. But such a matrix has two equal columns and therefore has a zero determinant, which we can see by applying Example 2.6.8 to the transpose of the matrix and using D7. So this sum must be 0 if $i \neq j$. We can combine these two sums by means of the Kronecker delta in the formula

$$\sum_{k=1}^{n} A_{ki} a_{kj} = \delta_{ij} \det A$$

In order to exploit this formula we make the following definitions:

Minor and Cofactor Matrices

DEFINITION 2.7.1. The *matrix of minors* of the $n \times n$ matrix $A = [a_{ij}]$ is the matrix $M(A) = [M_{ij}(A)]$ of the same size. The *matrix of cofactors* of A is the matrix $A_{cof} = [A_{ij}]$ of the same size. Finally, the *adjoint matrix* of A is the matrix $\operatorname{adj} A = A_{cof}^T$.

EXAMPLE 2.7.2. Compute the determinant, minors, cofactors and adjoint matrices for $A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & -1 \\ 0 & 2 & 1 \end{bmatrix}$ and compute $A \operatorname{adj} A$.

SOLUTION. The determinant is easily seen to be 2. Now for the matrix of minors:

$$
\begin{aligned}
M(A) &= \begin{bmatrix} \begin{vmatrix} 0 & -1 \\ 2 & 1 \end{vmatrix} & \begin{vmatrix} 0 & -1 \\ 0 & 1 \end{vmatrix} & \begin{vmatrix} 0 & 0 \\ 0 & 2 \end{vmatrix} \\ \begin{vmatrix} 2 & 0 \\ 2 & 1 \end{vmatrix} & \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 0 & 2 \end{vmatrix} \\ \begin{vmatrix} 2 & 0 \\ 0 & -1 \end{vmatrix} & \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 0 & 0 \end{vmatrix} \end{bmatrix} \\
&= \begin{bmatrix} 2 & 0 & 0 \\ 2 & 1 & 2 \\ -2 & -1 & 0 \end{bmatrix}
\end{aligned}
$$

To get the matrix of cofactors, simply overlay $M(A)$ with the following "checkerboard" of $+/-$'s

$$
\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}
$$

to obtain

$$
A_{cof} = \begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & -2 \\ -2 & 1 & 0 \end{bmatrix}
$$

Now transpose to get

$$
\operatorname{adj} A = \begin{bmatrix} 2 & -2 & -2 \\ 0 & 1 & 1 \\ 0 & -2 & 0 \end{bmatrix}
$$

We check that

$$
\begin{aligned}
(\operatorname{adj} A)A &= \begin{bmatrix} 2 & -2 & -2 \\ 0 & 1 & 1 \\ 0 & -2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & -1 \\ 0 & 2 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \\
&= (\det A) I_3
\end{aligned}
$$

Of course, the example simply confirms the formula that preceded it since this formula gives the (i, j)th entry of the product $(\operatorname{adj} A)A$. If we were to do determinants by row expansions, we would get a similar formula for the (i, j)th entry of $A \operatorname{adj} A$. We summarize this information in matrix notation as the determinantal property

Adjoint Formula

D8: For a square matrix A,

$$
A \operatorname{adj} A = (\operatorname{adj} A)A = (\det A)I
$$

What does this have to do with inverses? We already know that A is invertible exactly when $\det A \neq 0$, so the answer is staring at us! Just divide the terms in D8 by $\det A$ to obtain an explicit formula for A^{-1}:

Inverse Formula

D9: For a square matrix A such that $\det A \neq 0$,

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

EXAMPLE 2.7.3. Compute the inverse of the matrix A of Example 2.7.2 by the Inverse Formula.

SOLUTION. We already computed the adjoint matrix of A, and the determinant of A is just 2, so we have that

$$A^{-1} = \frac{1}{\det A}\operatorname{adj} A = \frac{1}{2}\begin{bmatrix} 2 & -2 & -2 \\ 0 & 1 & 1 \\ 0 & -2 & 0 \end{bmatrix}$$

EXAMPLE 2.7.4. Interpret the Inverse Formula in the case of the 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

SOLUTION. In this case we have

$$M(A) = \begin{bmatrix} d & c \\ b & a \end{bmatrix}$$
$$A_{cof} = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$$
$$\operatorname{adj} A = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

so that the Inverse Formula becomes

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

□

As you might expect, this is exactly the same as the formula we obtained in Example 2.5.8.

Cramer's Rule

Thanks to the Inverse formula, we can now find an explicit formula for solving linear systems with a nonsingular coefficient matrix. Here's how we proceed. To solve $A\mathbf{x} = \mathbf{b}$ we multiply both sides on the left by A^{-1} to obtain that $x = A^{-1}b$. Now use the Inverse formula to obtain

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\det A}\operatorname{adj}(A)\mathbf{b}$$

The explicit formula for the ith coordinate of $\mathbf{x}$ that comes from this fact is

$$x_i = \frac{1}{\det A}\sum_{j=1}^{n} A_{ji} b_j$$

The summation term is exactly what we would obtain if we started with the determinant of the matrix B_i obtained from A by replacing the ith column of A by $\mathbf{b}$ and then expanding the determinant down the ith column. Therefore, we have arrived at the following rule:

THEOREM 2.7.5. *Let A be an invertible $n \times n$ matrix and b an $n \times 1$ column vector. Denote by B_i the matrix obtained from A by replacing the ith column of A by b. Then the linear system $A\mathbf{x} = \mathbf{b}$ has unique solution $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ where*

$$x_i = \frac{\det B_i}{\det A}, \quad i = 1, 2, \ldots, n$$

Cramer's Rule

EXAMPLE 2.7.6. Use Cramer's rule to solve the system

$$\begin{aligned} 2x_1 - x_2 &= 1 \\ 4x_1 + 4x_2 &= 20 \end{aligned}$$

SOLUTION. The coefficient matrix and right hand side vectors are

$$A = \begin{bmatrix} 2 & -1 \\ 4 & 4 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 1 \\ 20 \end{bmatrix}$$

so that

$$\det A = 8 - (-4) = 12$$

and

$$x_1 = \frac{\begin{vmatrix} 1 & -1 \\ 20 & 4 \end{vmatrix}}{\begin{vmatrix} 2 & -1 \\ 4 & 4 \end{vmatrix}} = \frac{24}{12} = 2$$

$$x_2 = \frac{\begin{vmatrix} 2 & 1 \\ 4 & 20 \end{vmatrix}}{\begin{vmatrix} 2 & -1 \\ 4 & 4 \end{vmatrix}} = \frac{36}{12} = 3$$

□

Computational Efficiency of Determinants

The truth of the matter is that Cramer's Rule and adjoints are only good for small matrices and theoretical arguments. For if you evaluate determinants in a straightforward way from the definition, the work in doing so is about $2n!$ flops for an $n \times n$ system. (Recall that a "flop" in numerical linear algebra is a single addition or subtraction, or multiplication or division. For example, it is not hard to show that the operation of adding a multiple of one row vector of length n to another requires $2n$ flops. This number $2n!$ is vast when compared to the number $2n^3/3$ flops required for Gaussian elimination, even with "small" n, say $n = 10$. In this case we have $2 \cdot 10^3/3 \approx 667$, while $2 \cdot 10! = 7,527,600$.

On the other hand, there is a clever way to evaluate determinants that is much less work than the definition: use elementary row operations together with D2, D6 and the elementary operations that correspond to these rules to reduce the determinant to that of a triangular matrix. This will only require about $2n^3/3$ flops. As a matter of fact, it is tantamount to Gaussian elimination. But to use Cramer's Rule, you will have to calculate $n + 1$ determinants. So why bother with Cramer's Rule on larger problems when it still will take about n times as much work as Gaussian elimination? A similar remark applies to computing adjoints instead of using Gauss-Jordan elimination on the super-augmented matrix of A.

*Proofs of Some of the Laws of Determinants

D2: If B is obtained from A by multiplying one row of A by the scalar c, then $\det B = c \cdot \det A$.

To keep the notation simple, assume the first row is multiplied by c, the proof being similar for other rows. Suppose we have established this for all determinants of size less than n (this is really another "proof by induction", which is how most of the following determinantal properties are established). For an $n \times n$ determinant we have

$$\begin{aligned}\det B &= \begin{vmatrix} c \cdot a_{11} & c \cdot a_{12} & \cdots & c \cdot a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \\ &= c \cdot a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & & \vdots \\ a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} + \sum_{k=2}^{n} a_{k1}(-1)^{k+1} M_{k1}(B)\end{aligned}$$

But the minors $M_{k1}(B)$ all are smaller and have a common factor of c in the first row. Pull this factor out of every remaining term and we get that

$$\begin{vmatrix} c \cdot a_{11} & c \cdot a_{12} & \cdots & c \cdot a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = c \cdot \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

Thus we have shown that property D2 holds for all matrices.

D3: If B is obtained from A by interchanging two rows of A, then $\det B = -\det A$.

To keep the notation simple, assume we switch the first and second rows. In the case of a 2×2 determinant, we get the negative of the original determinant (check this for yourself). Suppose we have established the same is true for all matrices of size less than n. For an $n \times n$ determinant we have

$$\begin{aligned}\det B &= \begin{vmatrix} a_{21} & a_{22} & \cdots & a_{2n} \\ a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \\ &= a_{21} M_{11}(B) - a_{12} M_{21}(B) + \sum_{k=3}^{n} a_{k1}(-1)^{k+1} M_{k1}(B) \\ &= a_{21} M_{21}(A) - a_{12} M_{11}(A) + \sum_{k=3}^{n} a_{k1}(-1)^{k+1} M_{k1}(B)\end{aligned}$$

But all the determinants in the summation sign come from a submatrix of A with the first and second row interchanged. Since they are smaller than n, they are just the negative of the corresponding minor of A. Notice that the first two terms are just the first two terms in the determinantal expansion of A, except that they are out of order and have an extra minus sign. Factor this minus sign out of every term and we have obtained D3. □

D4: If B is obtained from A by adding a multiple of one row of A to another row of A, then $\det B = \det A$.

Actually, it's a little easier to answer a slightly more general question: what happens if we replace a row of a determinant by that row plus some other row vector **r** (not necessarily a row of the determinant)? Again, simply for convenience of notation, we assume the row in question is the first. The same argument works for any other row. Some notation: let B be the matrix which we obtain from the $n \times n$ matrix A by adding the row vector $\mathbf{r} = [r_1, r_2, \ldots, r_n]$ to the first row and C the matrix obtained from A by replacing the first row by **r**.The answer turns out to be that the $|B| = |A| + |C|$. One way of saying this is to say that the determinant function is "additive in each row." Let's see what happens in the one dimensional case:

$$|B| = |[a_{11} + r_1]| = a_{11} + r_1 = |[a_{11}]| + |[r_1]| = |A| + |C|$$

Suppose we have established the same is true for all matrices of size less than n and let A be $n \times n$. Then the minors $M_{k1}(B)$, with $k > 1$, are smaller than n so the property holds for them. Hence we have

$$\begin{aligned}
\det B &= \begin{vmatrix} a_{11} + r_1 & a_{12} + r_2 & \cdots & a_{1n} + r_n \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \\
&= (a_{11} + r_1) M_{11}(A) + \sum_{k=2}^{n} a_{k1} (-1)^{k+1} M_{k1}(B) \\
&= (a_{11} + r_1) M_{11}(A) + \sum_{k=2}^{n} a_{k1} (-1)^{k+1} (M_{k1}(A) + M_{k1}(C)) \\
&= \sum_{k=1}^{n} a_{k1} (-1)^{k+1} M_{k1}(A) + r_1 M_{11}(C) + \sum_{k=2}^{n} a_{k1} (-1)^{k+1} M_{k1}(C) \\
&= \det A + \det C
\end{aligned}$$

Now what about adding a multiple of one row to another in a determinant? For notational convenience, suppose we add s times the second row to the first. In the notation of the previous paragraph,

$$\det B = \begin{vmatrix} a_{11} + s \cdot a_{21} & a_{12} + s \cdot a_{22} & \cdots & a_{1n} + s \cdot a_{2n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

and

$$\det C = \begin{vmatrix} s \cdot a_{21} & s \cdot a_{22} & \cdots & s \cdot a_{2n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = s \cdot \begin{vmatrix} a_{21} & a_{22} & \cdots & a_{2n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = 0$$

where we applied D2 to pull the common factor s from the first row and the result of Example 2.6.8 to get the determinant with repeated rows to be 0. But $|B| = |A| + |C|$. Hence we have shown D4. □

D6: Given matrices A, B of the same size,

$$\det AB = \det A \det B$$

The key is that we now know that determinant calculation is intimately connected with elementary matrices, rank and the reduced row echelon form. First let's reinterpret D2-D4 still one more time. First of all take $A = I$ in the discussion of the previous paragraph and we see that

- $\det E_i(c) = c$
- $\det E_{ij} = -1$
- $\det E_{ij}(s) = 1$

Therefore, D2-D4 can be restated (yet again) as

- D2: $\det(E_i(c)A) = \det E_i(c) \cdot \det A$ (here $c \neq 0$.)
- D3: $\det(E_{ij}A) = \det E_{ij} \cdot \det A$
- D4: $\det(E_{ij}(s) = \det E_{ij}(s) \cdot \det A$

In summary: For any elementary matrix E and arbitrary matrix A of the same size, $\det(EA) = \det(E)\det(A)$.

Now let's consider this question: how does $\det(AB)$ relate to $\det(A)$ and $\det(B)$? If A is not invertible, $\operatorname{rank} A < n$ by Theorem 2.5.9 and so $\operatorname{rank} AB < n$ by Corollary 2.4.19. Therefore, $\det(AB) = 0 = \det A \cdot \det B$ in this case. Next suppose that A is invertible. Express it as a product of elementary matrices, say $A = E_1 E_2 \ldots E_k$ and use our summary of D1-D3 to disassemble and reassemble the elementary factors:

$$\begin{aligned}
\det(AB) &= \det(E_1 E_2 \ldots E_k B) \\
&= (\det E_1 \det E_2 \ldots \det E_k) \det B \\
&= \det(E_1 E_2 \ldots E_k) \det B \\
&= \det A \cdot \det B
\end{aligned}$$

Thus we have shown that **D6** holds. □

D7: For all square matrices A, $\det A^T = \det A$.

Recall these facts about elementary matrices:

- $\det E_{ij}^T = \det E_{ij}$
- $\det E_i(c)^T = \det E_i(c)$
- $\det E_{ij}(c)^T = \det E_{ji}(c) = 1 = \det E_{ij}(c)$

Therefore, transposing does not affect determinants of elementary matrices. Now for the general case observe that, since A and A^T are transposes of each other, one is invertible if and only if the other is by the Transpose/Inverse law. In particular, if both are singular, then $\det A^T = 0 = \det A$. On the other hand, if both are nonsingular, then write A as a product of elementary matrices, say $A = E_1 E_2 \ldots E_k$, and obtain from the product law for transposes that $A^T = E_k^T E_{k-1}^T \ldots E_1^T$, so by D6

$$\begin{aligned} \det A^T &= \det E_k^T \det E_{k-1}^T \ldots \det E_k^T \\ &= \det E_k \det E_{k-1} \ldots \det E_1 \\ &= \det E_1 \det E_2 \ldots \det E_k \\ &= \det A \end{aligned}$$

□

2.7 Exercises

1. For each of the following matrices find (1) the matrix of minors, (2) the matrix of cofactors, (3) the adjoint matrix for each matrix, and (4) the product of matrix and its adjoint.

(a) $\begin{bmatrix} 2 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & 2 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 3 \\ -1 & 2-i \end{bmatrix}$ (d) $\begin{bmatrix} -1 & 1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 2 \end{bmatrix}$

2. For each of the following matrices, find the inverses in two ways: first by superaugmented matrices, then by adjoints.

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

3. Use Cramer's Rule to solve the following systems.

(a) $\begin{aligned} x - 3y &= 2 \\ 2x + y &= 11 \end{aligned}$ (b) $\begin{aligned} 2x_1 + x_2 &= b_1 \\ 2x_1 - x_2 &= b_2 \end{aligned}$ (c) $\begin{aligned} 3x_1 + x_3 &= 2 \\ 2x_1 + 2x_2 &= 1 \\ x_1 + x_2 + x_3 &= 6 \end{aligned}$

4. Suppose we want to interpolate three points (x_k, y_k), $\quad k = 0, 1, 2$. Write out the system of equations that results from plugging these points into the equation of a quadratic $y = c_0 + c_1 x + c_2 x^2$ and calculate the determinant of the coefficient matrix. When is this determinant 0? (This coefficient matrix is an example of what is called a *Vandermonde* matrix.)

5. Confirm that the determinant of the matrix $A = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ is -1. We can now assert without any further calculation that the inverse matrix of A has integer coefficients, thanks to the adjoint formula. Explain.

6. Prove that if the matrix A is invertible, then $\operatorname{adj}(A^T A) > 0$.

7. Let A and B be invertible matrices of the same size. Prove the following.

(a) $\operatorname{adj} A^{-1} = (\operatorname{adj} A)^{-1}$

(b) $\operatorname{adj}(AB) = \operatorname{adj} A \ \operatorname{adj} B$

Hint: Determinantal law D9 can be very helpful here.

8. Suppose that the square matrix A is singular. Prove that if the system $A\mathbf{x} = \mathbf{b}$ is consistent, then $(\operatorname{adj} A)\,\mathbf{b} = \mathbf{0}$.

2.8. *Tensor Products

How do we solve a system of equations in which the unknowns can be organized into a matrix X and the linear system in question is of the form

$$AX - XB = C \tag{2.8.1}$$

where A, B, C are given matrices? We call this equation the *Sylvester equation*. Such systems occur in a number of physical applications; for example, discretizing certain partial differential equations in order to solve them numerically can lead to such a system. We are going to examine a matrix method for systematically reorganizing the data into a single column so that the resulting system looks like an ordinary linear system. The basic idea needed here is that of the tensor product of two matrices, which is defined as follows:

DEFINITION 2.8.1. Let $A = [a_{ij}]$ be an $m \times p$ matrix and $B = [b_{ij}]$ an $n \times q$ matrix. Then the *tensor product* of A and B is the $mn \times pq$ matrix which can be expressed in block form as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1j}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2j}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1}B & a_{i2}B & \cdots & a_{ij}B & \cdots & a_{in}B \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mj}B & \cdots & a_{mn}B \end{bmatrix}$$

EXAMPLE 2.8.2. Let $A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$ and describe the matrices $A \otimes B$, $B \otimes A$ and $I_2 \otimes A$ explicitly.

SOLUTION. First we have from the definition that

$$A \otimes B = \begin{bmatrix} 1B & 3B \\ 2B & 1B \end{bmatrix} = \begin{bmatrix} 4 & 12 \\ -1 & -3 \\ 8 & 4 \\ -2 & -1 \end{bmatrix}$$

and

$$B \otimes A = \begin{bmatrix} 4A \\ -1A \end{bmatrix} = \begin{bmatrix} 4 & 12 \\ -8 & -2 \\ -1 & -3 \\ -2 & -1 \end{bmatrix}$$

Similarly

$$I_2 \otimes A = \begin{bmatrix} 1A & 0A \\ 0A & 1A \end{bmatrix} = \begin{bmatrix} 1 & 3 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 2 & 1 \end{bmatrix}$$

□

We can think of the tensor product as a kind of matrix multiplication. One point that comes out of Example 2.8.2 is that, even though $A \otimes B$ and $B \otimes A$ have the same size, $A \otimes B \neq B \otimes A$ in general.

The other ingredient that we need to solve Equation 2.8.1 is an operator that turns matrices into vectors which is defined as follows.

DEFINITION 2.8.3. Let A be an $m \times n$ matrix. Then the $mn \times 1$ vector $\operatorname{vec} A$ is obtained from A by stacking the n columns of A vertically, with the first column at the top and the last column of A at the bottom.

EXAMPLE 2.8.4. Let $A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 4 \end{bmatrix}$. Compute $\operatorname{vec} A$.

SOLUTION. There are three columns to stack, yielding

$$\operatorname{vec} A = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \\ 2 \\ 4 \end{bmatrix}$$

□

Here are a few simple facts about tensor products that are more or less immediate from the definition.

THEOREM 2.8.5. *Let A, B, C, D be suitably sized matrices. Then*

1. $(A + B) \otimes C = A \otimes C + B \otimes C$
2. $A \otimes (B + C) = A \otimes B + A \otimes C$
3. $(A \otimes B) \otimes C = A \otimes (B \otimes C)$
4. $(A \otimes B)^T = A^T \otimes B^T$
5. $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$
6. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

The next theorem lays out the connection between tensor products and the vec operator.

THEOREM 2.8.6. *Let A, X, B be matrices conformable for multiplication. Then*

$$\operatorname{vec} AXB = \left(B^T \otimes A\right) \operatorname{vec} X$$

The proof of this statement amounts to checking corresponding entries of each side of the equation above; we leave this to the reader. It is easy to check that the vec operator is linear, that is, $\operatorname{vec}(A + B) = \operatorname{vec} A + \operatorname{vec} B$. As a consequence, we have this very useful fact, which we state for just two summands.

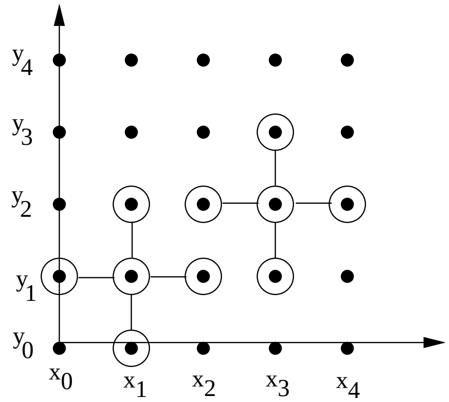


FIGURE 2.8.1. Molecules for $(1,1)$th and $(3,2)$th grid points.

COROLLARY 2.8.7. *Any solution matrix X to the linear system*

$$A_1 X B_1 + A_2 X B_2 = C$$

satisfies the linear system

$$\left(\left(B_1^T \otimes A_1\right) + \left(B_2^T \otimes A_2\right)\right) \operatorname{vec} X = \operatorname{vec} C$$

The following is a very basic application of the tensor product. Suppose we wish to model a two dimensional heat diffusion process on a flat plate that occupies the unit square in the xy-plane. We proceed as we did in the one dimensional process described in the introduction of Chapter 1. To fix ideas, we assume that the heat source is described by a function $f(x,y), 0 \le x \le 1, 0 \le y \le 1$, and that the temperature is held at 0 at the boundary of the unit square. Also, the conductivity coefficient is assumed to be the constant k. Cover the square with a uniformly spaced set of grid points $(x_i, y_j), 0 \le i,j \le n+1$, called nodes, and assume that the spacing in each direction is a width $h = 1/(n+1)$. Also assume that the temperature function at the (i,j)th node is $u_{ij} = u(x_i, y_j)$ and that the source is $f_{ij} = f(x_i, y_j)$. Notice that the values of u on boundary grid points is set at 0. For example, $u_{01} = u_{20} = 0$. By balancing the heat flow in the horizontal and vertical directions, one arrives at a system of linear equations, one for each node, of the form

$$-u_{i-1,j} - u_{i+1,j} + 4u_{ij} - u_{i,j-1} - u_{i,j+1} = \frac{h^2}{k} f_{ij}, \quad i,j = 1, \ldots, n \tag{2.8.2}$$

Observe that values of boundary nodes are zero, so these are not unknowns, which is why the indexing of the equations starts at 1 instead of 0. There are exactly as many equations as unknown grid point values. Each equation has a “molecule” associated with it which is obtained by circling the nodes that occur in the equation and connecting these circles. A picture of a few nodes is given in Figure 2.8.1.

EXAMPLE 2.8.8. Set up and solve a system of equations for the two dimensional heat diffusion problem described above.

SOLUTION. Equation 2.8.2 gives us n^2 equations in the n^2 unknowns u_{ij}, $i,j = 1,2,\dots,n$. Rewrite Equation 2.8.2 in the form

$$(-u_{i-1,j} + 2u_{ij} - u_{i+1,j}) + (-u_{i,j-1} + 2u_{ij} - u_{i,j+1}) = \frac{h^2}{k} f_{ij}$$

Now form the $n \times n$ matrices

$$T_n = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & \ddots & 0 \\ 0 & \ddots & \ddots & -1 \\ 0 & 0 & 2 & 2 \end{bmatrix}$$

$U = [u_{ij}]$ and $F = [f_{ij}]$ and we see that the general equations can be written in matrix form as

$$T_n U + U T_n = T_n U I_n + I_n U T_n = \frac{h^2}{k} F.$$

However, we can't as yet identify a coefficient matrix, which is where Corollary 2.8.7 comes in handy. Note that both I_n and T_n are symmetric and apply the Corollary to obtain that the system has the form

$$(I_n \otimes T_n + T_n \otimes I_n) \operatorname{vec} U = \operatorname{vec} \frac{h^2}{k} F.$$

Now we have a coefficient matrix and, what's more, we have an automatic ordering of the doubly indexed variables u_{ij}, namely

$$u_{1,1}, u_{2,1}, \dots, u_{n,1}, u_{1,2}, u_{2,2}, \dots, u_{n,2}, \dots, u_{1,n}, u_{2,n}, \dots, u_{n,n}.$$

This is sometimes called the "row ordering," which refers to the rows of the nodes in Figure 2.8.1, and not the rows of the matrix U. □

2.8 Exercises

1. Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}$. Write the matrices $A \otimes B$ and $B \otimes A$ explicitly.

2. With A, B as above, $C = \begin{bmatrix} 2 & -1 \\ 1 & 0 \\ 1 & 3 \end{bmatrix}$, and $X = [x_{ij}]$ a 3×2 matrix of unknowns, use tensor products to determine the coefficient matrix of the linear system $AX + XB = C$.

3. Verify parts 1 and 4 of Theorem 2.8.5.

4. Verify parts 5 and 6 of Theorem 2.8.5.

5. If heat is transported with a horizontal velocity v as well as diffused in Example 2.8.8 a new equation results at each node in the form

$$-u_{i-1,j} - u_{i+1,j} + 4u_{ij} - u_{i,j-1} - u_{i,j+1} - \frac{vh}{2k}(u_{i+1,j} - u_{i-1,j}) = \frac{h^2}{k} f_{ij}$$

for $i, j = 1, \dots, n$. Vectorize the system and use tensor products to identify the coefficient matrix of this linear system.

2.9. *Computational Notes and Projects

LU Factorization

Here is a problem: suppose we want to solve a nonsingular linear system $Ax = b$ repeatedly, with different choices of b. A perfect example of this kind of situation is the heat flow problem Example 1.1.5 where the right hand side is determined by the heat source term $f(x)$. Suppose that we need to experiment with different source terms. What happens if we do straight Gaussian elimination or Gauss-Jordan elimination? Each time we carry out a complete calculation on the augmented matrix $\widetilde{A} = [A \,|\, b]$ we have to resolve the whole system. Yet, the main part of our work is the same: putting the part of $\widetilde{A}$ corresponding to the coefficient matrix A into reduced row echelon form. Changing the right hand side has no effect on this work. What we want here is a way to somehow record our work on A, so that solving a new system involves very little additional work. This is exactly what the LU factorization is all about.

DEFINITION 2.9.1. Let A be an $n \times n$ matrix. An LU factorization of A is a pair of $n \times n$ matrices L, U such that

1. L is lower triangular.
2. U is upper triangular.
3. $A = LU$.

Even if we could find such beasts, what is so wonderful about them? The answer is that *triangular* systems $Ax = b$ are easy to solve. For example, if A is upper triangular, we learned that the smart thing to do was to use the last equation to solve for the last variable, then the next to the last equation for the next to the last variable, etc. This is the secret of Gaussian elimination! But lower triangular systems are just as simple: use the first equation to solve for the first variable, the second equation for the second variable, and so forth. Now suppose we want to solve $Ax = b$ and we know that $A = LU$. The original system becomes $LUx = b$. Introduce an intermediate variable $y = Ux$. Now perform these steps:

1. (Forward solve) Solve lower triangular system $Ly = b$ for the variable y.
2. (Back solve) Solve upper triangular system $Ux = y$ for the variable x.

This does it! Once we have the matrices L, U, we don't have to worry about right hand sides, except for the small amount of work involved in solving two triangular systems. Notice, by the way, that since A is assumed nonsingular, we have that if $A = LU$, then

$\det A = \det L \det U \neq 0$. Therefore, neither triangular matrix L or U can have zeros on its diagonal. Thus, the forward and back solve steps can always be carried out to give a unique solution.

EXAMPLE 2.9.2. You are given that

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

Use this fact to solve $Ax = b$, where (a) $b = [1, 0, 1]^T$ and (b) $b = [-1, 2, 1]^T$.

SOLUTION. Set $x = [x_1, x_2, x_3]^T$ and $y = [y_1, y_2, y_3]$. For (a) forward solve

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

to get $y_1 = 1$, then $y_2 = 0 + 1y_1 = 1$, then $y_3 = 1 - 1y_1 - 2y_2 = -2$. Then back solve

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

to get $x_3 = -2/(-1) = 2$, then $x_2 = 1 + x_3 = 3$, then $x_1 = (1 - 1x_2)/2 = -1$.

For (b) forward solve

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

to get $y_1 = -1$, then $y_2 = 0 + 1y_1 = -1$, then $y_3 = 1 - 1y_1 - 2y_2 = 4$. Then back solve

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix}$$

to get $x_3 = 4/(-1) = -4$, then $x_2 = 1 + x_3 = -3$, then $x_1 = (1 - 1x_2)/2 = 2$. □

Notice how simple the previous example was, given the LU factorization. Now how do we find such a factorization? In general, a nonsingular matrix may not have such a factorization. A good example is the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. However, if Gaussian elimination can be performed on the matrix A *without row exchanges*, then such a factorization is really a by-product of GE. In this case let $[a_{ij}^{(k)}]$ be the matrix obtained from A after using the kth pivot to clear out entries below it (thus $A = [a_{ij}^{(0)}]$). Remember that in GE we only need two types of elementary operations, namely row exchanges and adding a multiple of one row to another. Furthermore, the only elementary operations of the latter type that we use are of this form: $E_{ij}(-a_{jj}^{(k)}/a_{ij}^{(k)})$, where $[a_{ij}^{(k)}]$ is the matrix obtained from A from the various elementary operations up to this point. The numbers $m_{ij} = -a_{jj}^{(k)}/a_{ij}^{(k)}$, where $i > j$, are sometimes called *multipliers*. In the way of notation, let us call a triangular matrix a *unit* triangular matrix if its diagonal entries are all 1's.

THEOREM 2.9.3. *If Gaussian elimination is used without row exchanges on the nonsingular matrix A, resulting in the upper triangular matrix U, and if L is the unit lower triangular matrix whose entries below the diagonal are the negatives of the multipliers m_{ij}, then $A = LU$.*

PROOF. The proof of this theorem amounts to noticing that the product of all the elementary operations that reduces A to U is a unit lower triangular matrix $\widetilde{L}$ with the multipliers m_{ij} in the appropriate positions. Thus $\widetilde{L}A = U$. To undo these operations, multiply by a matrix L with the negatives of the multipliers in the appropriate positions. This results in

$$L\widetilde{L}A = A = LU$$

as desired. □

The following example shows how one can write an efficient program to implement LU factorization. The idea is this: as we do Gaussian elimination the U part of the factorization gradually appears in the upper parts of the transformed matrices $A^{(k)}$. Below the diagonal we replace nonzero entries with zeros, column by column. Instead of wasting this space, use it to store the negative of the multipliers in place of the element it zeros out. Of course, this storage part of the matrix should not be changed by subsequent elementary row operations. When we are finished with elimination, the diagonal and upper part of the resulting matrix is just U and the strictly lower triangular part on the unit lower triangular matrix L is stored in the lower part of the matrix.

EXAMPLE 2.9.4. Use the shorthand of the preceding discussion to compute an LU factorization for

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

SOLUTION. Proceed as in Gaussian elimination, but store negative multipliers:

$$\begin{bmatrix} 2 & 1 & 0 \\ -2 & 0 & -1 \\ 2 & 3 & -3 \end{bmatrix} \xrightarrow[E_{31}(-1)]{E_{21}(1)} \begin{bmatrix} 2 & 1 & 0 \\ -1 & 1 & -1 \\ 1 & 2 & -3 \end{bmatrix} \xrightarrow[E_{32}(-2)]{} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ -1 & 2 & -1 \end{bmatrix}$$

Now we read off the results from the last matrix:

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix} \text{ and } U = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & -1 \end{bmatrix}$$ □

What can be said if pivoting is required (for example, we might want to use a partial pivoting strategy)? Take the point of view that we could see our way to the end of Gaussian elimination and store the product P of all row exchanging elementary operations that we use along the way. A product of such matrices is called a *permutation matrix*; such a matrix is invertible, since it is a product of invertible matrices. Thus if we apply the correct permutation matrix P to A we obtain a matrix for which Gaussian elimination will succeed without further row exchanges. Consequently, we have a theorem that applies to all nonsingular matrices. Notice that it does not limit the usefulness of LU factorization since the linear system $Ax = b$ is equivalent to the system $PAx = Pb$. The following theorem could be called the "PLU factorization theorem."

THEOREM 2.9.5. *If A is a nonsingular matrix, then there exists a permutation matrix P, upper triangular matrix U, and unit lower triangular matrix L such that $PA = LU$.*

There are many other useful factorizations of matrices that numerical analysts have studied, e.g., LDU and Cholesky. We will stop at LU, but there is one last point we want to make. The amount of work in finding the LU factorization is the same as Gaussian elimination itself, which we saw in Section 1.5 of Chapter 1 is approximately $2n^3/3$ flops. The addition work of back and forward solving is about $2n^2$ flops. So the dominant amount of work is done by computing the factorization rather than the back and forward solving stages.

Project Topics

Project: LU Factorization

Write a program module that implements Theorem 2.9.5 using partial pivoting and implicit row exchanges. This means that space is allocated for the $n \times n$ matrix $A = [a[i, j]]$ and an array of row indices, say $indx[i]$. Initially, $indx$ should consist of the integers $1, 2, \ldots, n$. Whenever two rows need to be exchanged, say e.g., the first and third, then the indices $indx[1]$ and $indx[3]$ are exchanged. References to array elements throughout the Gaussian elimination process should be indirect: refer to the $(1, 4)$th entry of A as the element $a[indx[1], 4]$. This method of reference has the same effect as physically exchanging rows, but without the work. It also has the appealing feature that we can design the algorithm as though no row exchanges have taken place provided we replace the direct reference $a[i, j]$ by the indirect reference $a[indx[i], j]$. The module should return the lower/upper matrix in the format of Example 2.9.4 as well as the permuted array $indx[i]$. Effectively, this index array tells the user what the permutation matrix P is.

Next write an LU system solver module that uses the LU factorization to solve a general linear system.

Finally, write a module that finds the inverse of an $n \times n$ matrix A by first using the LU factorization module, then making repeated use of the LU system solver to solve $A\mathbf{x}^{(i)} = \mathbf{e}_i$, where $\mathbf{e}_i$ is the ith column of the identity. Then we will have

$$A^{-1} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}]$$

Be sure to document and test your code. Report on the results of its application.

Project: Markov Chains

Refer to Example 2.3.4 and Section 2.3 for background. Three automobile insurance firms compete for a fixed market of customers. Annual premiums are sold to these customers. We will label the companies A, B and C. You work for Company A, and your team of market analysts has done a survey which draws the following conclusions: in each of the past three years, the number of A customers switching to B is 20%, and to C is 30%. The number of B customers switching to A is 20%, and to C is 20%. The number of C customers switching to A is 30%, and to B is 10%. Those who do not switch continue to use their current company's insurance for the next year. The first part of your problem is to model this market as a Markov chain. Display the transition matrix for the model. To illustrate the workings of the model, show what it would predict as the market shares three years from now if currently A, B and C owned equal shares of the market.

The next part of your problem is as follows: your team has tested two advertising campaigns in some smaller test markets and are confident that the first campaign will convince 20% of the B customers who would otherwise stay with B in a given year to switch to A. The second advertising campaign would convince 20% of the C customers who would otherwise stay with C in a given year to switch to A. Both campaigns have about equal costs and would not change other customers habits. You have to make a recommendation, based on your experiments with various possible initial state vectors for the market. Will these campaigns actually improve your company's market share? If so, which one are you going to recommend to your superiors? Write up your recommendation in the form of a report, with supporting evidence. It's a good idea to hedge on your bets a little by pointing out limitations to your model and claims, so devote a few sentences to those points.

It would be a plus to carry the analysis further (your manager might appreciate that). For instance, you could turn the additional market share from, say B customers, into a variable and plot the long term gain for your company against this variable. A manager could use this data to decide if it were worthwhile to attempt gaining more customers from B. This is a bit open ended and optional.

Project: Modeling with Directed Graphs I

Refer to Example 2.3.7 and Section 2.3 for background. As a social scientist you have studied the influence factors that relate seven coalition groups which, for simplicity, we will simply label $1, 2, 3, 4, 5, 6, 7$. Based on empirical studies, you conclude that the influence factors can be well modeled by a dominance-directed graph with each group as a vertex. The meaning of the presence of an edge (i, j) in the graph is that coalition group i can dominate, i.e., swing coalition group j its way on a given political issue. The data you have gathered suggests that the appropriate edge set is the following:

$$\begin{aligned} E &= \{(1,2),(1,3),(1,4),(1,7),(2,4),(2,6),(3,2),(3,5),(3,6),\\ &\quad (4,5),(4,7),(5,1),(5,6),(5,7),(6,1),(6,4),(7,2),(7,6)\} \end{aligned}$$

Do an analysis of this power structure. This should include a graph. (It might be a good idea to arrange the vertices in a circle and go from there.) It should also include a power rating of each coalition group. Now suppose you were an advisor to one of these coalition groups and, by currying certain favors, this group could gain influence over another coalition group (thereby adding an edge to the graph or reversing an existing edge of the graph). In each case, if you could pick the best group for your client to influence, which would that be? Explain your results in the context of matrix multiplication if you can.

2.9 Exercises

1. Find the LU factorization of $A = \begin{bmatrix} 2 & -1 & 1 \\ 2 & 3 & -2 \\ 4 & 2 & -2 \end{bmatrix}$ and use it to solve the system $A\mathbf{x} = \mathbf{b}$ where $\mathbf{b}$ is

(a) $[6, -8, -4]^T$ (b) $(2, -1, 2)$ (c) $(1, 2, 4)$

2. Show that if A is a nonsingular matrix with a zero $(1, 1)$th entry, then A does not have an LU factorization.

3. Find a PLU factorization of $A = \begin{bmatrix} 0 & -1 & 1 \\ 2 & 3 & -2 \\ 4 & 2 & -2 \end{bmatrix}$, and use it to solve the system $A\mathbf{x} = \mathbf{b}$ where $\mathbf{b}$ is

(a) $(3, 1, 4)$ (b) $(2, -1, 3)$ (c) $(1, 2, 0)$

Review

Chapter 2 Exercises

1. Let $A = \begin{bmatrix} 2 & -1 & 1 \\ 2 & 3 & -2 \\ 4 & 2 & -2 \end{bmatrix}$ and $\mathbf{x} = [x, y, z]$. Then the equation $(\mathbf{x}A)^T + A\mathbf{x}^T = [1, 4, 2]^T$ represents a linear system in the variables x, y, z. Find the coefficient matrix of this system.

2. Determine for what values of k the matrix $A = \begin{bmatrix} 2 & 1 \\ k & 3 \end{bmatrix}$ is invertible and find the inverse in that case.

3. Find the determinant of $A = \begin{bmatrix} 2 & 1 & 0 & 3 \\ 0 & 2 & -1 & 0 \\ 0 & -1 & 2 & 0 \\ -1 & 0 & -1 & 2 \end{bmatrix}$.

4. Show by example that the sum of invertible matrices need not be invertible.

5. Show that if A is any square matrix, then $A + A^T$ is symmetric. Use this to show that every quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ can be defined by a symmetric matrix $B = (A + A^T)/2$ as well. Apply this result to the matrix of Example 2.4.16.

6. A square matrix A is called *normal* if $A^H A = AA^H$. Determine which of the following matrices are normal:

$$\text{(a)} \begin{bmatrix} 2 & i \\ 1 & 2 \end{bmatrix}, \text{(b)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}, \text{(c)} \begin{bmatrix} 1 & i \\ 1 & 2+i \end{bmatrix}$$

7. Express the matrix $D = \begin{bmatrix} 3 & 3 \\ 1 & -3 \end{bmatrix}$ as a linear combination of the matrices $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ and $C = \begin{bmatrix} 0 & 2 \\ 0 & -1 \end{bmatrix}$.

8. Find all possible products of two matrices from among the following:

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 4 \end{bmatrix}, \quad C = \begin{bmatrix} 1 \\ 5 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 3 & 0 \\ -1 & 2 & 1 \end{bmatrix}$$

9. Prove that if $D = ABC$, where A, C and D are invertible matrices, then B is invertible.

10. Use a block multiplication to find the square of $\begin{bmatrix} 3 & 1 & -1 \\ 2 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

11. Given that $C = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ in block form with A and B square, show that C is invertible if and only if A and B are, in which case

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

12. Show by example that a sum or product of nilpotent matrices need not be nilpotent.

13. Suppose that $A = B + C$, where B is a symmetric matrix and C is a skew-symmetric matrix. Show that $B = \frac{1}{2}(A + A^T)$ and $B = \frac{1}{2}(A - A^T)$.

14. Let T be an upper triangular matrix.

(a) Show that $T = D + M$, where D is diagonal and M is strictly upper triangular.

(b) If D is invertible, show that $T = D(I - N)$, where N is strictly upper triangular.

(c) If D is invertible, use (b) and Exercise 14 to obtain a formula for T^{-1} involving D and N.

CHAPTER 3

VECTOR SPACES

It is hard to overstate the importance of the idea of a vector space, a concept which has found application in the areas of mathematics, engineering, physics, chemistry, biology, the social sciences and others. What we encounter is an abstraction of the idea of vector space that we studied in calculus. In this Chapter, abstraction will come in two waves. The first wave, which could properly be called generalization, consists of generalizing the familiar ideas of geometrical vectors of calculus to vectors of size greater than three. These vector spaces could still be regarded as "concrete." The second wave consists of abstracting the vector idea to entirely different kinds of objects. Abstraction can sometimes be difficult. For some, the study of abstract ideas is its own reward. For others, the natural reaction is to expect some payoff for the extra effort required to master abstraction. In the case of vector spaces we are happy to report that both kinds of students will be satisfied: vector space theory really is a thing of beauty in itself and there is indeed a payoff for its study. It is a practical tool that enables us to understand phenomena that would otherwise escape our comprehension. For example, in this chapter we will use the theory in network analysis and in finding the "best" solution to an inconsistent system (least squares), as well as new perspectives on our old friend $A\mathbf{x} = \mathbf{b}$.

3.1. Definitions and Basic Concepts

Generalization

We begin with the most concrete form of vector spaces, one that is closely in tune with what we learned in calculus, when we were first introduced to two and three dimensional vectors. Bear in mind that in calculus we were only concerned with real numbers as scalars. However, we have seen that the complex numbers are a perfectly legitimate (and sometimes more useful than the reals) field of numbers to work with. Therefore, our concept of a vector space must include the selection of a field of scalars. The requirements for such a field are that it have binary operations of addition and multiplication which satisfy the usual arithmetic laws: both operations are closed, commutative, associative, have identities, satisfy distributive laws, and that there exist additive inverses and multiplicative inverses for nonzero elements. Although other fields are possible, for our purposes the only fields of scalars are $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. As has been the case previously in this text, unless there is some indication to the contrary, the field of scalars will be assumed to be the default, the real numbers $\mathbb{R}$.

A formal definition of vector space will come later. For now we describe a "vector space" over a field of scalars $\mathbb{F}$ as a nonempty set V of vectors of the same size, together with the binary operations of scalar multiplication and vector addition, subject to the following laws: for all vectors $\mathbf{u}, \mathbf{v} \in V$ and scalars $a \in \mathbb{F}$, (a) (Closure of vector addition) $\mathbf{u} + \mathbf{v} \in V$. (b) (Closure of scalar multiplication) $a\mathbf{v} \in V$. Concrete Vector Spaces

Very simple examples are $\mathbb{R}^2$ and $\mathbb{R}^3$ which we discuss below. Another is any line through the origin in $\mathbb{R}^2$, which takes the form $V = \{c(x_0, y_0) \mid c \in \mathbb{R}\}$.

NOTATION 3.1.1. For vectors $\mathbf{u}, \mathbf{v}$, we define $-\mathbf{u} = (-1)\mathbf{u}$ and $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v})$.

Geometrical vector spaces. We have already seen the vector idea in geometry or calculus. In those contexts, a vector was supposed to represent a direction and a magnitude in two or three dimensional space. At first, one had to deal with these intuitive definitions until they could be turned into something more explicitly computational, namely the following two vector spaces over the field of real numbers:

$$\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$$
$$\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$$

The distinction between vector spaces and ordinary geometrical spaces becomes a little hazy here. Once we have set up a coordinate system we can identify each point in two or three dimensional space with its coordinates, which we write in the form of a tuple, i.e., a vector. The arithmetic of these two vector spaces are just the usual coordinate-wise vector addition and scalar multiplication. One can visualize the direction represented by a vector (x, y) by drawing an arrow, i.e., directed line segment, from the origin (the point with coordinates $(0, 0)$ in the plane) to the point with coordinates (x, y). The magnitude of this vector is the length of the arrow, which is just $\sqrt{x^2 + y^2}$. The arrows that we draw only *represent* the vector we are thinking of. More than one arrow could represent the same vector as in Figure 3.3.1. The definitions of vector arithmetic could be represented geometrically too. For example, to get the sum of vectors $\mathbf{u}$ and $\mathbf{v}$, one places a representative of vector $\mathbf{u}$ in the plane, then places a representative of $\mathbf{v}$ whose tail is at the head of $\mathbf{v}$, and the vector $\mathbf{u} + \mathbf{v}$ is then represented by the third leg of this triangle, with base at the base of $\mathbf{u}$. To get a scalar multiple of a vector $\mathbf{w}$ one scales $\mathbf{w}$ in accordance with the coefficient. See Figure 3.1.1.Though instructive, this version of vector addition is not practical for calculations.

As a practical matter, it is also convenient to draw directed line segments connecting points; such a vector is called a displacement vector. For example, see Figure 3.1.1 for representatives of a displacement vector $\mathbf{w} = \overrightarrow{PQ}$ from the point P with coordinates $(1, 2)$ to the point Q with coordinates $(3, 3)$. One of the first nice outcomes of vector arithmetic is that this displacement vector can be deduced from a simple calculation Displacement Vector

$$\mathbf{w} = (3, 3) - (1, 2) = (3 - 1, 3 - 2) = (2, 1)$$

As a matter of fact, this example has familiar objects in it. We already agreed in Chapter 2 to use the tuple notation as a shorthand for column vectors. The arithmetic of $\mathbb{R}^2$ and $\mathbb{R}^3$ is the same as the usual arithmetic for column vectors. Now even though we can't draw real geometrical pictures of vectors with four or more coordinates, we have seen that larger vectors are useful in our search for solutions of linear systems. So the question presents itself: why stop at three? The answer is that we won't! We will

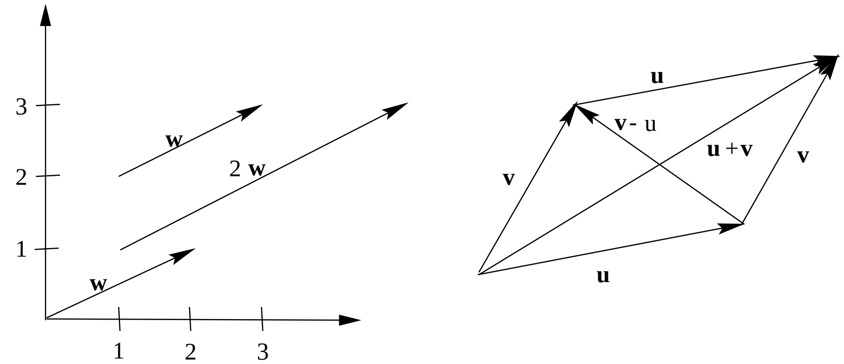


FIGURE 3.1.1. Displacement vectors and graphical vector operations.

use the familiar pictures of $\mathbb{R}^2$ and $\mathbb{R}^3$ to guide our intuition about vectors in higher dimensional spaces, which we now present.

Standard Vector Spaces

DEFINITION 3.1.2. Given a positive integer n, we define the *standard vector space of dimension n over the reals* to be the set of vectors

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) \mid x_1, x_2, \dots, x_n \in \mathbb{R}\}$$

together with the usual vector addition and scalar multiplication. (Remember that $(x_1, x_2, \dots, x_n)$ is shorthand for the column vector $[x_1, x_2, \dots, x_n]^T$.)

We see immediately from definition that the required closure properties of vector addition and scalar multiplication hold, so these really are vector spaces in the sense defined above. The standard real vector spaces are often called the real Euclidean vector spaces once the notion of a norm (a notion of length covered in the next section) is attached to them. As in Chapter 2, we don't have to stop at the reals. For those situations in which we want to use complex numbers, we have the following vector spaces:

DEFINITION 3.1.3. Given a positive integer n, we define the *standard vector space of dimension n over the complex numbers* to be the set of vectors

$$\mathbb{C}^n = \{(x_1, x_2, \dots, x_n) \mid x_1, x_2, \dots, x_n \in \mathbb{C}\}$$

together with the usual vector addition and scalar multiplication.

The standard complex vector spaces are also sometimes called Euclidean spaces. It's rather difficult to draw honest spatial pictures of complex vectors. The space $\mathbb{C}^1$ isn't too bad: complex numbers can be identified by points in the complex plane. What about $\mathbb{C}^2$? Where can we put $(1 + 2i, 3 - i)$? It seems like we need four real coordinates, namely the real and imaginary parts of two independent complex numbers, to keep track of. This is too big to fit in real three dimensional space, where we have only three independent coordinates. We don't let this technicality deter us. We can still draw fake vector pictures of elements of $\mathbb{C}^2$ to help our intuition, but do the algebra of vectors exactly from definition.

EXAMPLE 3.1.4. Find the displacement vector from the point P with coordinates $(1 + 2i, 1 - 2i)$ to the point Q with coordinates $(3 + i, 2i)$.

SOLUTION. We compute

$$\begin{aligned}\overrightarrow{PQ} &= (3+i, 2i) - (1+2i, 1-2i) \\ &= (3+i-(1+2i), 2i-(1-2i)) \\ &= (2-i, -1+4i)\end{aligned}$$

□

Abstraction

Now we examine the abstraction of our concept of vector space. First we have to identify the essential vector spaces properties, enough to make the resulting structure rich, but not so much that it is tied down to an overly specific form. We saw in Chapter 2 that many laws hold for the standard vector spaces. The essential laws were summarized in Section 2.1 of Chapter 2. These laws become the basis for our definition of an abstract vector space.

Abstract Vector Space

DEFINITION 3.1.5. An *(abstract) vector space* is a nonempty set V of elements called vectors, together with operations of vector addition ($+$) and scalar multiplication ($\cdot$), such that the following laws hold: for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and scalars $a, b \in \mathbb{F}$,

1. (Closure of vector addition) $\mathbf{u} + \mathbf{v} \in V$.
2. (Commutativity of addition) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
3. (Associativity of addition) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
4. (Additive identity) There exists an element $0 \in V$ such that $\mathbf{u} + 0 = \mathbf{u} = 0 + \mathbf{u}$.
5. (Additive inverse) There exists an element $-\mathbf{u} \in V$ such that $\mathbf{u} + (-\mathbf{u}) = 0 = (-\mathbf{u}) + \mathbf{u}$.
6. (Closure of scalar multiplication) $a\mathbf{u} \in V$.
7. (Distributive law) $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.
8. (Distributive law) $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$.
9. (Associative law) $(ab)\mathbf{u} = a(b\mathbf{u})$.
10. (Monoidal law) $1\mathbf{u} = \mathbf{u}$.

Examples of these abstract vector spaces are the standard spaces just introduced, and these will be our main focus in this section. Yet, if we squint a bit, we can see vector spaces everywhere. There are other, entirely non-standard examples, which make the abstraction worthwhile. Here are just a few such of examples. Our first example is closely related to the standard spaces, though strictly speaking it is not one of them. It blurs the distinction between matrices and vectors in Chapter 2, since it makes matrices into "vectors" in the abstract sense of the preceding definition.

EXAMPLE 3.1.6. Let $\mathbb{R}^{m,n}$ denote the set of all $m \times n$ matrices with real entries. Show this set, with the usual matrix addition and scalar multiplication, forms a vector space.

SOLUTION. We know that any two matrices of the same size can be added to yield a matrix of that size. Likewise, a scalar times a matrix yields a matrix of the same size. Thus the operations of matrix addition and scalar multiplication are closed. Indeed, these laws and all the other vector space laws are summarized in the laws of matrix addition and scalar multiplication of page 53. □

The next example is important in many areas of higher mathematics and is quite different from the standard vector spaces, yet all the same a perfectly legitimate vector space. All the same, at first it feels odd to think of functions as "vectors" even though this is meant in the abstract sense.

EXAMPLE 3.1.7. Let $C[0,1]$ denote the set of all real valued functions that are continuous on the interval $[0,1]$ and use the standard function addition and scalar multiplication for these functions. That is, for $f(x), g(x) \in C[0,1]$ and real number c, we define the functions $f+g$ and cf by

$$(f+g)(x) = f(x) + g(x)$$
$$(cf)(x) = c(f(x)).$$

Show that $C[0,1]$ with the given operations is a vector space.

SOLUTION. We set $V = C[0,1]$ and check the vector space axioms for this V. For the rest of this example, we let f, g, h be arbitrary elements of V. We know from calculus that the sum of any two continuous functions is continuous and that any constant times a continuous function is also continuous. Therefore the closure of addition and scalar multiplication hold. Now for all x such that $0 \leq x \leq 1$, we have from definition and the commutative law of real number addition that

$$(f+g)(x) = f(x) + g(x) = g(x) + f(x) = (g+f)(x)$$

Since this holds for all x, we conclude that $f+g = g+f$, which is the commutative law of vector addition. Similarly,

$$\begin{aligned}((f+g)+h)(x) &= (f+g)(x) + h(x) = (f(x)+g(x)) + h(x) \\ &= f(x) + (g(x)+h(x)) = (f+(g+h))(x)\end{aligned}$$

Since this holds for all x, we conclude that $(f+g)+h = f+(g+h)$, which is the associative law for addition of vectors.

Next, if 0 denotes the constant function with value 0, then for any $f \in V$ we have that for all $0 \leq x \leq 1$,

$$(f+0)(x) = f(x) + 0 = f(x)$$

Since this is true for all x we have that $f+0 = f$, which establishes the additive identity law. Also, we define $(-f)(x) = -(f(x))$ so that for all $0 \leq x \leq 1$,

$$(f+(-f))(x) = f(x) - f(x) = 0$$

from which we see that $f + (-f) = 0$. The additive inverse law follows. For the distributive laws note that for real numbers a, b and continuous functions $f, g \in V$, we have that for all $0 \leq x \leq 1$,

$$a(f+g)(x) = a(f(x)+g(x)) = af(x) + ag(x) = (af+ag)(x)$$

which proves the first distributive law. For the second distributive law, note that for all $0 \leq x \leq 1$,

$$((a+b)g)(x) = (a+b)g(x) = ag(x) + bg(x) = (ag+bg)(x)$$

and the second distributive law follows. For the scalar associative law, observe that for all $0 \leq x \leq 1$,

$$((ab)f)(x) = (ab)f(x) = a(bf(x)) = (a(bf))(x)$$

so that $(ab)f = a(bf)$, as required. Finally, we see that

$$(1f)(x) = 1f(x) = f(x)$$

from which we have the monoidal law $1f = f$. Thus, $C[0,1]$ with the prescribed operations is a vector space. $\square$

The preceding example could have just as well been $C[a,b]$, the set of all continuous functions on the interval $a \leq x \leq b$. Indeed, most of what we say about $C[0,1]$ is equally applicable to the more general space $C[a,b]$. We stick to the interval $0 \leq x \leq 1$ for simplicity. The next example is also based on the "functions as vectors" idea.

EXAMPLE 3.1.8. Let $V = \{f(x) \in C[0,1] \mid f(1/2) = 0\}$ and $W = \{f(x) \in C[0,1] \mid f(1/2) = 1\}$, where each set has the operations of function addition and scalar multiplication as in Example 3.1.7. One of these sets forms a vector space over the reals, while the other does not. Determine which.

SOLUTION. Notice that we don't have to check the commutativity of addition, associativity of addition, distributive laws, associative law or monoidal law. The reason is that we already know from the previous example that these laws hold when the operations of the space $C[0,1]$ are applied to any elements of $C[0,1]$, whether they belong to V or W or not. So the only laws to be checked are the closure laws and the identity laws.

First let $f(x), g(x) \in V$ and let c be a scalar. By definition of the set V we have that $f(1/2) = 0$ and $g(1/2) = 0$. Add these equations together and we obtain

$$(f+g)(1/2) = f(1/2) + g(1/2) = 0 + 0 = 0$$

It follows that V is closed under addition with these operations. Furthermore, if we multiply the identity $f(1/2) = 0$ by the real number c we obtain that

$$(cf)(1/2) = c \cdot f(1/2) = c \cdot 0 = 0$$

It follows that V is closed under scalar multiplication. Now certainly the zero function belongs to V, since this function has value 0 at any argument. Therefore, V contains an additive identity element. Finally, we observe that the negative of a function $f(x) \in V$ is also an element of V, since

$$(-f)(1/2) = -1 \cdot f(1/2) = -1 \cdot 0 = 0$$

Therefore, the set V with the given operations satisfies all the vector space laws and is an (abstract) vector space in its own right.

When we examine the set W in a similar fashion, we run into a roadblock at the closure of addition law. If $f(x), g(x) \in W$, then by definition of the set W we have that $f(1/2) = 1$ and $g(1/2) = 1$. Add these equations together and we obtain

$$(f+g)(1/2) = f(1/2) + g(1/2) = 1 + 1 = 2$$

This means that $f+g$ is not in W, so the closure of addition fails. We need go no further. If only one of the vector space axioms fails, then we do not have a vector space. Hence, W with the given operations is not a vector space. $\square$

Notice that there is a certain economy in this situation. A number of laws did not need to be checked by virtue of the fact that the sets in question were subsets of existing vector spaces with the same vector operations. We shall have more to say about this situation in the next section. Here is another example that is useful and instructive.

EXAMPLE 3.1.9. Show that the set S_n of all $n \times n$ real symmetric matrices with the usual matrix addition and scalar multiplication form a vector space.

SOLUTION. Just as in the preceding example, we don't have to check the commutativity of addition, associativity of addition, distributive laws, associative law or monoidal law since we know that these laws hold for any matrices, symmetric or not. Now let $A, B \in S_n$. This means by definition that $A = A^T$ and $B = B^T$. Let c be any scalar. Then we have both

$$(A + B)^T = A^T + B^T = A + B$$

and

$$(cA)^T = cA^T = cA$$

It follows that the set S_n is closed under the operations of matrix addition and scalar multiplication. Furthermore, the zero $n \times n$ matrix is clearly symmetric, so the set S_n has an additive identity element. Finally, $(-A)^T = -A^T = -A$, so each element of S_n has an additive inverse as well. Therefore, all of the laws for a vector space are satisfied, so S_n together with these operations is an (abstract) vector space. □

One of the virtues of abstraction is that it allows us to cover many cases with one statement. For example, there are many simple facts that are deducible from the vector space laws alone. With the standard vector spaces, these facts seem transparently clear. For abstract spaces, the situation is not quite so obvious. Here are a few examples of what can be deduced from definition.

EXAMPLE 3.1.10. Let $\mathbf{v} \in V$, a vector space and 0 the vector zero. Deduce *from the vector space properties alone* that $0\mathbf{v} = 0$.

SOLUTION. Certainly we have the scalar identity $0 + 0 = 0$. Multiply both sides on the right by the vector $\mathbf{v}$ to obtain that

$$(0 + 0)\mathbf{v} = 0\mathbf{v}$$

Now use the distributive law to obtain

$$0\mathbf{v} + 0\mathbf{v} = 0\mathbf{v}$$

Next add $-(0\mathbf{v})$ to both sides (remember, we don't know it's 0 yet), use the associative law of addition to regroup and obtain that

$$0\mathbf{v} + (0\mathbf{v} + (-0\mathbf{v})) = 0\mathbf{v} + (-0\mathbf{v})$$

Now use the additive inverse law to obtain that

$$0\mathbf{v} + 0 = 0$$

Finally, use the identity law to obtain

$$0\mathbf{v} = 0$$

which is what we wanted to show. □

EXAMPLE 3.1.11. Show that the vector space V has only one zero element.

SOLUTION. Suppose that both 0 and 0_* act as zero elements in the vector space. Use the additive identity property of 0 to obtain that $0_* + 0 = 0_*$, while the additive identity property of 0_* implies that $0 + 0_* = 0$. By the commutative law of addition, $0_* + 0 = 0 + 0_*$. It follows that $0_* = 0$, whence there can be only one zero element. □

There are several other such arithmetic facts that we want to identify, along with the one of this example. In the case of standard vectors, these facts are obvious, but for abstract vector spaces, they require a proof similar to the one we have just given. We leave these as exercises.

Laws of Vector Arithmetic. Let $\mathbf{v}$ be a vector in some vector space V and let c be any scalar. Then

1. $0\mathbf{v} = \mathbf{0}$
2. $c\mathbf{0} = \mathbf{0}$
3. $(-c)\mathbf{v} = c(-\mathbf{v}) = -(c\mathbf{v})$
4. If $c\mathbf{v} = \mathbf{0}$, then $\mathbf{v} = \mathbf{0}$ or $c = 0$.
5. A vector space has only one zero element.
6. Every vector has only one additive inverse.

A reminder about notation: just as in matrix arithmetic, for vectors $\mathbf{u}, \mathbf{v} \in V$, we understand that $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v})$.

Linear Operators

We were introduced in Section 2.3 of Chapter 2 to the idea of a linear function in the context of standard vectors. Now that we have a notion of an abstract vector space, we can examine linearity in this larger setting. We have seen that some of our "vectors" can themselves be functions, as in the case of the vector space $C[0, 1]$ of continuous functions on the interval $[0, 1]$. In order to avoid confusion in cases like this, we prefer to designate linear functions by the term *linear operator.* Other common terms for this object are *linear mapping* or *linear transformation.*

Before giving the definition of linear operator, let us recall some notation that is associated with functions in general. We identify a function f with the notation $f : D \to T$, where D and T are the *domain* and *target* of the function, respectively. This means that for each x in the domain D, the value $f(x)$ is a uniquely determined element in the target T. We want to emphasize at the outset, that there is a difference here between the *target* of a function and its *range*. The *range* of the function f is defined as the subset of the target

$$\operatorname{range}(f) = \{y \mid y = f(x) \text{ for some } x \in D\}$$

which is just the set of all possible values of $f(x)$. For example, we can define a function $f : \mathbb{R} \to \mathbb{R}$ by the formula $f(x) = x^2$. It follows from our specification of f that the target of f is understood to be $\mathbb{R}$, while the range of f is the set of nonnegative real numbers.

A function that maps elements of one vector space into another, say $f : V \to W$ is sometimes called an *operator* or *transformation*. For example, the operator $f : \mathbb{R}^2 \to$

$\mathbb{R}^3$ might be given by the formula

$$f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x^2 \\ xy \\ y^2 \end{bmatrix}$$

Notice in this example that the target of f is $\mathbb{R}^3$, which is not the same as the range of f, since elements in the range have nonnegative first and third coordinates. From the point of view of linear algebra, this function lacks the essential feature that makes it really interesting, namely linearity.

DEFINITION 3.1.12. A function $T : V \to W$ from the vector space V into the space W over the same field of scalars is called a *linear operator (mapping, transformation)* if, for all vectors $\mathbf{u}, \mathbf{v} \in V$ and scalars c, d, we have

$$T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v})$$

By taking $c = d = 1$ in the definition, we see that a linear function T is *additive,* that is, $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$. Also, by taking $d = 0$ in the definition, we see that a linear function is *outative,* that is, $T(c\mathbf{u}) = cT(\mathbf{u})$. As a matter of fact, these two conditions imply the linearity property, and so are equivalent to it. We leave this fact as an exercise.

By repeated application of the linearity definition, we can extend the linearity property to any linear combination of vectors, not just two terms. This means that for any scalars $c_1, c_2, \ldots, c_n$ and vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, we have

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_nT(\mathbf{v}_n)$$

EXAMPLE 3.1.13. Determine if $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear operator, where T is given by the formula (a) $T((x, y)) = (x^2, xy, y^2)$ or (b) $T((x, y)) = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$.

SOLUTION. If T is defined by (a) then we show by a simple example that T fails to be linear. Let us calculate

$$T((1, 0) + (0, 1)) = T((1, 1)) = (1, 1, 1)$$

while

$$T((1, 0)) + T((0, 1)) = (1, 0, 0) + (0, 0, 1) = (1, 0, 1).$$

These two are not equal, so T fails to satisfy the linearity property.

Next consider the operator T defined as in (b). Write $A = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ and we see that the action of T can be given as $T(\mathbf{v}) = A\mathbf{v}$. Now we have already seen in Section 2.3 of Chapter 2 that the operation of multiplication by a fixed matrix is a linear operator. $\square$

In Chapter 2 the following useful fact was shown, which we now restate for real vectors, though it is equally valid for standard complex vectors.

THEOREM 3.1.14. *Let A be an $m \times n$ matrix and define an operator $T_A : \mathbb{R}^n \to \mathbb{R}^m$ by the formula $T(\mathbf{v}) = A\mathbf{v}$, for all $\mathbf{v} \in \mathbb{R}^n$. Then T_A is a linear operator.*

Abstraction gives us a nice framework for certain key properties of mathematical objects, some of which we have seen before. For example, in calculus we were taught that differentiation has the "linearity property." Now we can view this assertion in a fuller context: let V be the space of differentiable functions and define an operator T on V by the rule $T(f(x)) = f'(x)$. Then T is a linear operator on the space V.

3.1 Exercises

In Exercises 1-6 you are to determine if the given set and operations define a vector space. If not, indicate which laws fail.

1. $V = \{\begin{bmatrix} a & b \\ 0 & a+b \end{bmatrix} | a, b \in \mathbb{R}\}$ with the usual matrix addition and scalar multiplication.

2. $V = \{\begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} | a \in \mathbb{R}\}$ with the usual matrix addition and scalar multiplication.

3. $V = \{\begin{bmatrix} a & b & \overline{a} \end{bmatrix} | a, b \in \mathbb{C}\}$ with the usual matrix addition and scalar multiplication. In this example the scalar field is the complex numbers.

4. V consists of all continuous functions $f(x)$ on the interval $[0, 1]$ such that $f(0) = 0$.

5. V consists of all quadratic polynomial functions $f(x) = ax^2 + bx + c, a \neq 0$.

6. V consists of all continuous functions $f(x)$ on the interval $[0, 1]$ such that $f(0) = f(1)$.

7. Use the definition of vector space to prove Vector Law of Arithmetic 2: $c\mathbf{0} = \mathbf{0}$.

8. Use the definition of vector space to prove Vector Law of Arithmetic 3: $(-c)\mathbf{v} = c(-\mathbf{v}) = -(c\mathbf{v})$.

9. Use the definition of vector space to prove Vector Law of Arithmetic 4: If $c\mathbf{v} = \mathbf{0}$, then $\mathbf{v} = \mathbf{0}$ or $c = 0$.

10. Let $\mathbf{u}, \mathbf{v} \in V$, where V is a vector space. Use the vector space laws to prove that the equation $\mathbf{x} + \mathbf{u} = \mathbf{v}$ has one and only one solution vector $\mathbf{x} \in V$, namely $\mathbf{x} = \mathbf{u} - \mathbf{v}$.

11. Let V be a vector space and form the set V^2 consisting of all ordered pairs $(\mathbf{u}, \mathbf{v})$ where $\mathbf{u}, \mathbf{v} \in V$. We can define an addition and scalar multiplication on these ordered pairs as follows

$$(\mathbf{u}_1, \mathbf{v}_1) + (\mathbf{u}_1, \mathbf{v}_1) \equiv (\mathbf{u}_1 + \mathbf{u}_2, \mathbf{v}_1 + \mathbf{v}_2)$$
$$c \cdot (\mathbf{u}_1, \mathbf{v}_1) = (c\mathbf{u}_1, c\mathbf{v}_1)$$

Verify that with these operations V^2 becomes a vector space over the same field of scalars as V.

12. Determine which of the following functions $T : \mathbb{R}^3 \to \mathbb{R}^2$ is a linear operator and if so, write the operator as a matrix multiplication. Here $\mathbf{x} = (x, y, z)$

(a) $T(\mathbf{x}) = (x - y, x + 2y - 4z)$ (b) $T(\mathbf{x}) = (x + y, xy)$ (c) $T(\mathbf{x}) = (y, z, x)$

13. Let $V = C[0, 1]$ and define an operator $T : V \to V$ by the following formulas for $T(f)$:

(a) $T(f)(x) = f(1)x^2$ (b) $T(f)(x) = f^2(x)$ (c) $T(f)(x) = 2f(x)$ (d) $T(f)(x) = \int_0^x f(s)\, ds$

Which, if any of these operators is linear? If so, is the target V of the operator equal to its range?

14. Determine if the operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation or not (give reasons) where

(a) $T(x, y) = x(0, y)$ (b) $T(x, y) = (x + 2y, 0)$ (c) $T(x, y) = (\sin x, \cos y)$

15. Let $T : \mathbb{R}^3 \to \mathcal{P}_2$ be defined by $T((a, b, c)) = a + bx + cx^2$. Show that T is a linear operator whose range is $\mathcal{P}_2$.

16. Prove the remark following Definition 3.1.12: if a function $T : V \to W$ between vector spaces V and W is additive and outative, then it is linear.

3.2. Subspaces

We now turn our attention to the concept of a *subspace*, which is a rich source of useful examples of vector spaces. It frequently happens that a certain vector space of interest is a subset of a larger, and possibly better understood vector space, and that the vector operations are the same for both spaces. A good example of this situation is given by the vector space V of Example 3.1.8 which is a subset of the larger vector space $C[0, 1]$ with both spaces sharing the same definitions of vector addition and scalar multiplication. Here is a precise definition for the subspace idea.

DEFINITION 3.2.1. A *subspace* of the vector space V is a subset W of V such that W, together with the binary operations it inherits from V, forms a vector space (over the same field of scalars as V) in its own right.

Given a subset W of the vector space V, we can apply the definition of vector space directly to the subset W to obtain the following very useful test.

Subspace Test

THEOREM 3.2.2. *Let W be a subset of the vector space V. Then W is a subspace of V if and only if*

1. *W contains the zero element of V.*
2. *(Closure of addition) For all $\mathbf{u}, \mathbf{v} \in W$, $\mathbf{u} + \mathbf{v} \in W$.*
3. *(Closure of scalar multiplication) For all $\mathbf{u} \in W$ and scalars c, $c\mathbf{u} \in W$.*

PROOF. Let W be a subspace of the vector space V. Then the closure of addition and scalar multiplication are automatically satisfied by the definition of vector space. For condition 1, we note that W must contain a zero element by definition of vector space. Let 0^* be this element and 0 the zero element of V, so that $0^* + 0^* = 0^*$. Add the negative of 0^* (as an element of V) to both sides, cancel terms and we see that $0^* = 0$. This shows that W satisfies condition 1.

Conversely, suppose that W is a subset of V satisfying the three conditions. Since the operations of W are those of the vector space V, and V is a vector space, most of the laws for W are automatic. Specifically, the laws of commutativity, associativity, distributivity and the monoidal law hold for elements of W. The additive identity law follows from condition 1.

The only law that needs any work is the additive inverse law. Let $\mathbf{w} \in W$. By closure of scalar multiplication, $(-1)\mathbf{w}$ is in W. By the laws of vector arithmetic in the preceding section, this vector is simply $-\mathbf{w}$. This proves that every element of W has an additive inverse in W, which shows that W is a subspace of V. □

One notable point that comes out of the subspace test is that every subspace of V contains the zero vector. This is certainly not true of arbitrary subsets of V and serves to remind us that, although every subspace is a subset of V, not every subset is a subspace. Confusing the two is a common mistake, so much so that we issue the following

Caution: Every subspace of a vector space is a subset, but not conversely.

EXAMPLE 3.2.3. Which of the following subsets of the standard vector space $V = \mathbb{R}^3$ are subspaces of V?

(a) $W_1 = \{(x, y, z) \mid x - 2y + z = 0\}$ (b) $W_2 = \{(x, y, z) \mid x, y, \text{and } z \text{ are positive}\}$
(c) $W_3 = \{(0, 0, 0)\}$ (d) $W_4 = \{(x, y, z) \mid x^2 - y = 0\}$

SOLUTION. (a) Take $\mathbf{w} = (0, 0, 0)$ and obtain that

$$0 - 2 \cdot 0 + 0 = 0,$$

so that $\mathbf{w} \in W_1$. Next, check closure of W_1 under addition. Let's name two general elements from W_1, say $\mathbf{u} = (x_1, y_1, z_1)$ and $\mathbf{v} = (x_2, y_2, z_2)$. Then we know from definition of W_1 that

$$\begin{aligned} x_1 - 2y_1 + z_1 &= 0 \\ x_2 - 2y_2 + z_2 &= 0 \end{aligned}$$

We want to show that $\mathbf{u} + \mathbf{v} = (x_1 + x_2, y_1 + y_2, z_1 + z_2) \in W_1$. So add the two equations above and group terms to obtain

$$(x_1 + x_2) - 2(y_1 + y_2) + (z_1 + z_2) = 0.$$

This equation shows that the coordinates of $\mathbf{u} + \mathbf{v}$ fit the requirement for being an element of W_1, i.e., $\mathbf{u} + \mathbf{v} \in W_1$. Similarly, if c is a scalar then we can multiply the equation that says $\mathbf{u} \in W_1$, i.e., $x_1 - 2y_1 + z_1 = 0$, by c to obtain

$$(cx_1) - 2(cy_1) + (cz_1) = c0 = 0.$$

This shows that the coordinates of $c\mathbf{v}$ fit the requirement for being in W_1, i.e., $c\mathbf{u} \in W_1$. It follows that W_1 is closed under both addition and scalar multiplication, so it is a subspace of $\mathbb{R}^3$.

(b) This one is easy. Any subspace must contain the zero vector $(0, 0, 0)$. Clearly W_2 does not. Hence it cannot be a subspace. Another way to see it is to notice that closure under scalar multiplication fails (try multiplying $(1, 1, 1)$ by -1).

(c) The only possible choice for arbitrary elements $\mathbf{u}, \mathbf{v}$, in this space are $\mathbf{u} = \mathbf{v} = (0, 0, 0)$. But then we see that W_3 obviously contains the zero vector and for any scalar c

$$
\begin{aligned}
(0, 0, 0) + (0, 0, 0) &= (0, 0, 0) \\
c(0, 0, 0) &= (0, 0, 0)
\end{aligned}
$$

Therefore W_3 is a subspace of V by the subspace test.

(d) First of all, $0^2 - 0 = 0$, which means that $(0, 0, 0) \in W_4$. Likewise we see that $(1, 1, 0) \in W_4$ as well. But $(1, 1, 0) + (1, 1, 0) = (2, 2, 0)$ which is not an element of W_4 since $2^2 - 2 \neq 0$. Therefore, closure of addition fails and W_4 is not a subspace of V by the subspace test. □

Part (c) of this example highlights part of a simple fact about vector spaces. Every vector space V must have at least two subspaces, namely, $\{0\}$ (where 0 is the zero vector in V) and V itself. These are not terribly exciting subspaces, so they are commonly called the *trivial* subspaces.

EXAMPLE 3.2.4. Show that the subset $\mathcal{P}$ of $C[0, 1]$ consisting of all polynomial functions is a subspace of $C[0, 1]$ and that the subset $\mathcal{P}_n$ consisting of all polynomials of degree at most n is a subspace of $\mathcal{P}$.

SOLUTION. Certainly $\mathcal{P}$ is a subset of $C[0, 1]$, since any polynomial is uniquely determined by its values on the interval $[0, 1]$ and $\mathcal{P}$ contains the zero constant function which is a polynomial function. Let f and g be two polynomial functions on the interval $[0, 1]$, say

$$
\begin{aligned}
f(x) &= a_0 + a_1 x + \cdots + a_n x^n \\
g(x) &= b_0 + b_1 x + \cdots + b_n x^n
\end{aligned}
$$

where n is an integer equal to the maximum of the degrees of $f(x)$ and $g(x)$. Let c be any real number and we see that

$$
\begin{aligned}
(f + g)(x) &= (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n \\
(cf)(x) &= ca_0 + ca_1 x + \cdots + ca_n x^n
\end{aligned}
$$

which shows that $\mathcal{P}$ is closed under function addition and scalar multiplication. By the subspace test $\mathcal{P}$ is a subspace of $C[0, 1]$. The very same equations above also show that the subset $\mathcal{P}_n$ passes the subspace test, so it is a subspace of $C[0, 1]$. □

EXAMPLE 3.2.5. Show that the set of all upper triangular matrices (see page 74) in the vector space $V = \mathbb{R}^{n,n}$ of $n \times n$ real matrices is a subspace of V.

SOLUTION. Since the zero matrix is upper triangular, the subset W of all upper triangular matrices contains the zero element of V. Let $A = [a_{i,j}]$ and $B = [b_{i,j}]$ be any two matrices in W and let c be any scalar. By definition of upper triangular, we must have $a_{i,j} = 0$ and $b_{i,j} = 0$ if $i > j$. However,

$$\begin{aligned} A + B &= [a_{i,j} + b_{i,j}] \\ cA &= [ca_{i,j}]. \end{aligned}$$

and for $i > j$ we have $a_{i,j} + b_{i,j} = 0 + 0 = 0$ and $ca_{i,j} = c0 = 0$, so that $A + B$ and cA are also upper triangular. It follows that W is a subspace of V by the subspace test. □

There is an extremely useful type of subspace which requires the notion of a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ in the vector space V: an expression of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

where $c_1, c_2, \dots, c_n$ are scalars. We can consider the set of all possible linear combinations of a given list of vectors, which is what our next definition is about.

DEFINITION 3.2.6. Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ be vectors in the vector space V. The *span* of these vectors, denoted by $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, is the subset of V consisting of all possible linear combinations of these vectors, i.e.,

Linear Combinations and Span

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n \mid c_1, c_2, \dots, c_n \text{are scalars}\}$$

Caution: The scalars we are using really make a difference. For example, if $\mathbf{v}_1 = (1, 0)$ and $\mathbf{v}_2 = (0, 1)$ are viewed as elements of the real vector space $\mathbb{R}^2$, then

$$\begin{aligned} \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} &= \{c_1(1,0) + c_2(0,1) \mid c_1, c_2 \in \mathbb{R}\} \\ &= \{(c_1, c_2) \mid c_1, c_2 \in \mathbb{R}\} \\ &= \mathbb{R}^2 \end{aligned}$$

Similarly, if we view $\mathbf{v}_1$ and $\mathbf{v}_2$ as elements of the complex vector space $\mathbb{C}^2$, then we see that $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} = \mathbb{C}^2$. Now $\mathbb{R}^2$ consists of those elements of $\mathbb{C}^2$ whose coordinates have zero imaginary parts, so $\mathbb{R}^2$ is a sub*set* of $\mathbb{C}^2$; but these are certainly not equal sets. By the way, $\mathbb{R}^2$ is definitely not a sub*space* of $\mathbb{C}^2$ either, since the subset $\mathbb{R}^2$ is not closed under multiplication by complex scalars.

We should take note here that the definition of span would work perfectly well with infinite sets, as long as we understand that linear combinations in the definition would be finite and therefore not involve all the vectors in the span. A situation in which this extension is needed is as follows: consider the space $\mathcal{P}$ of all polynomials with the usual addition and scalar multiplication. It makes perfectly good sense to write

$$\mathcal{P} = \operatorname{span}\{1, x, x^2, x^3, \dots, x^n, \dots\}$$

since every polynomial is a *finite* linear combination of various monomials x^k.

EXAMPLE 3.2.7. Interpret the following linear spans in $\mathbb{R}^3$ geometrically:

$$\text{(a) } W_1 = \operatorname{span}\left\{\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}\right\}, \quad \text{(b) } W_2 = \operatorname{span}\left\{\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}\right\}$$

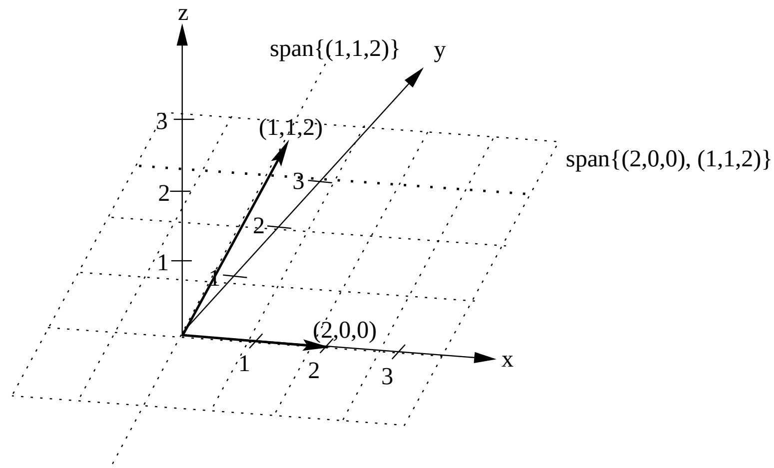


FIGURE 3.2.1. Cross-hatched portion of span{(2,0,0), (1,1,2)} and span{(1,1,2)}.

SOLUTION. Elements of W_1 are simply scalar multiples of the single vector $(1, 1, 2)$. The set of all such multiples gives us a line through the origin $(0, 0, 0)$. On the other hand, elements of W_2 give all possible linear combinations of two vectors $(1, 1, 2)$ and $(2, 0, 0)$. The locus of points generated by these combinations is a plane in $\mathbb{R}^3$ containing the origin, so it is determined by the points with coordinates $(0, 0, 0)$, $(1, 1, 2)$, and $(2, 0, 0)$. See Figure 3.2.1 for a picture of these spans. □

Spans are the premier examples of subspaces. In a certain sense, it can be said that *every* subspace is the span of some of its vectors. The following important fact is a very nice application of the subspace test.

THEOREM 3.2.8. *Let* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *be vectors in the vector space* V. *Then* $W = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *is a subspace of* V.

PROOF. First, we observe that W the zero vector can be expressed as the linear combination $0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_n$, which is an element of W. Next, let c be any scalar and form general elements $\mathbf{u}, \mathbf{v} \in W$, say

$$\begin{aligned}\mathbf{u} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n \\ \mathbf{v} &= d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_n\mathbf{v}_n\end{aligned}$$

Add these vectors and collect like terms to obtain

$$\mathbf{u} + \mathbf{v} = (c_1 + d_1)\mathbf{v}_1 + (c_2 + d_2)\mathbf{v}_2 + \cdots + (c_n + d_n)\mathbf{v}_n$$

Thus $\mathbf{u} + \mathbf{v}$ is also a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, so W is closed under vector addition. Finally, form the product $c\mathbf{u}$ to obtain

$$c\mathbf{u} = (cc_1)\mathbf{v}_1 + (cc_2)\mathbf{v}_2 + \cdots + (cc_n)\mathbf{v}_n$$

which is again a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, so W is closed under scalar multiplication. By the subspace test, W is a subspace of V. □

There are a number of simple properties of spans that we will need from time to time. One of the most useful is this basic fact.

THEOREM 3.2.9. *Let* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *be vectors in the vector space* V *and let* $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ *be vectors in* $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. *Then*

$$\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\} \subseteq \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

PROOF. Suppose that for each index $j = 1, 2, \ldots k$,

$$\mathbf{w}_j = c_{1j}\mathbf{v}_1 + c_{2j}\mathbf{v}_2 + \cdots + c_{nj}\mathbf{v}_n.$$

Then we may write a linear combination of the $\mathbf{w}_j$'s by regrouping the coefficients of each $\mathbf{v}_k$ as

$$\begin{aligned} d_1\mathbf{w}_1 + d_2\mathbf{w}_2 + \cdots + d_k\mathbf{w}_k &= d_1(c_{11}\mathbf{v}_1 + c_{21}\mathbf{v}_2 + \cdots + c_{n1}\mathbf{v}_n) \\ &\quad + d_2(c_{12}\mathbf{v}_1 + c_{22}\mathbf{v}_2 + \cdots + c_{n2}\mathbf{v}_n) + \ldots \\ &\quad + d_k(c_{1k}\mathbf{v}_1 + c_{2k}\mathbf{v}_2 + \cdots + c_{nk}\mathbf{v}_n) \\ &= \left(\sum_{i=1}^{k} d_i c_{i1}\right)\mathbf{v_1} + \left(\sum_{i=1}^{k} d_i c_{i1}\right)\mathbf{v_2} + \ldots \left(\sum_{i=1}^{k} d_i c_{i1}\right)\mathbf{v_n}. \end{aligned}$$

It follows that each element of $\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ belongs to the vector space $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, as desired. □

Here is a simple application of this theorem: if $\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, \ldots, \mathbf{v}_{i_k}$ is a subset of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, then

$$\operatorname{span}\{\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, \ldots, \mathbf{v}_{i_k}\} \subseteq \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

The reason is that each

$$\mathbf{w}_j = \mathbf{v}_{i_j} = 0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 1\mathbf{v}_{i_j} + \cdots + 0\mathbf{v}_n$$

so that the theorem applies to these vectors. Put another way, if we enlarge the list of spanning vectors, we enlarge the spanning set. However, we may not obtain a strictly larger spanning set, as the following example shows.

EXAMPLE 3.2.10. Show that

$$\operatorname{span}\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\} = \operatorname{span}\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}\}$$

Why might one prefer the first spanning set?

SOLUTION. Label vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Every element of $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ belongs to $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, since we can write $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + 0\mathbf{v}_3$. So we certainly have that $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} \subseteq \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. However, a little fiddling with numbers reveals this fact:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} = (-1)\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

In other words $\mathbf{v}_3 = -\mathbf{v}_1 + 2\mathbf{v}_2$. Therefore any linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ can be written as

$$\begin{aligned} c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3(-\mathbf{v}_1 + 2\mathbf{v}_2) \\ &= (c_1 - c_3)\mathbf{v}_1 + (c_2 + 2c_3)\mathbf{v}_2 \end{aligned}$$

Thus any element of $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ belongs to $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$, so the two spans are equal. This is an algebraic representation of the geometric fact that the three vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ belong to the same plane in $\mathbb{R}^2$ that is spanned by the two vectors $\mathbf{v}_1, \mathbf{v}_2$. It seems reasonable that we should prefer the spanning set $\mathbf{v}_1, \mathbf{v}_2$ to the set $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, since the former is smaller yet carries just as much information as the latter. As a matter of fact, we would get the same span if we used $\mathbf{v}_1, \mathbf{v}_3$ or $\mathbf{v}_2, \mathbf{v}_3$. The spanning set $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ has “redundant” vectors in it. $\square$

3.2 Exercises

In Exercises 1-7, determine if the subset W is a subspace of the given vector space V.

1. $V = \mathbb{R}^3$ and $W = \{(a, b, a - b + 1) \mid a, b \in \mathbb{R}\}$.

2. $V = \mathbb{R}^3$ and $W = \{(a, 0, a - b) \mid a, b \in \mathbb{R}\}$.

3. $V = \mathbb{R}^3$ and $W = \{(a, b, c) \mid 2a - b + c = 0\}$.

4. $V = \mathbb{R}^{2,3}$ and $W = \{\begin{bmatrix} a & b & 0 \\ b & a & 0 \end{bmatrix} \mid a, b \in \mathbb{R}\}$.

5. $V = C[0,1]$ and $W = \{f(x) \in C[0,1] \mid f(1) + f(1/2) = 0\}$.

6. $V = C[0,1]$ and $W = \{f(x) \in C[0,1] \mid f(1) \leq 0\}$.

7. $V = \mathbb{R}^{n,n}$ and W is the set of all invertible matrices in V.

8. Recall that a matrix A is *skew-symmetric* if $A^T = -A$.

(a) Show that every skew-symmetric matrix has the form $A = \begin{bmatrix} a & b \\ -b & c \end{bmatrix}$, for some scalars a, b, c.

(b) Show that the set V of all 2×2 skew-symmetric real matrices is a subspace of $\mathbb{R}^{2,2}$.

9. Show that $\operatorname{span}\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\} = \mathbb{R}^2$.

10. Show that $\operatorname{span}\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\} = \mathbb{R}^2$.

11. Which of the following spans equal the space $\mathcal{P}_2$ of polynomials of degree at most 2? Justify your answers.

(a) $\operatorname{span}\{1, 1 + x, x^2\}$ (b) $\operatorname{span}\{x, 4x - 2x^2, x^2\}$ (c) $\operatorname{span}\{1 + x + x^2, 1 + x, 3\}$

12. Find two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ such that if $\mathbf{u} = (1, -1, 1)$, then $\mathbb{R}^3 = \operatorname{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$

13. Let $\mathbf{u} = (2, -1, 1)$, $\mathbf{v} = (0, 1, 1)$ and $\mathbf{w} = (2, 1, 3)$. Show that $\operatorname{span}\{\mathbf{u} + \mathbf{w}, \mathbf{v} - \mathbf{w}\} \subseteq \operatorname{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ and determine whether or not these spans are actually equal.

14. Prove that if $V = \mathbb{R}^{n,n}$, then the set of all diagonal matrices is a subspace of V.

15. Let U and V be subspaces of W. Use the subspace test to prove the following.

(a) The set intersection $U \cap V$ is a subspace of W.

(b) The sum of the spaces, $U + V = \{u + v \mid u \in U \text{ and } v \in V\}$ is a subspace of W.

(c) The set union $U \cup V$ is not a subspace of W unless one of U or V is contained in the other.

16. Let V and W be subspaces of $\mathbb{R}^3$ given by

$$V = \{(x, y, z) \mid x = y = z \in \mathbb{R}\} \text{and } W = \{(x, y, 0) \mid x, y \in \mathbb{R}\}$$

Show that $V + W = \mathbb{R}^3$ and $V \cap W = \{\mathbf{0}\}$.

17. Let V be the space of 2×2 matrices and associate with each $A \in V$ the vector $\operatorname{vec}(A) \in \mathbb{R}^4$ obtained from A by stacking the columns of A underneath each other. (For example, $\operatorname{vec}(\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}) = (1, -1, 2, 1)$.) Prove the following

(a) The vec operation establishes a one-to-one correspondence between matrices in V and vectors in $\mathbb{R}^4$.

(b) The vec operation preserves operations in the sense that if A, B are matrices and c, d scalars, then

$$\operatorname{vec}(cA + dB) = c \operatorname{vec}(A) + d \operatorname{vec}(B)$$

18. You will need a computer algebra system (CAS) such as Mathematica or Maple for this exercise. Use the matrix

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

and the translation method of the preceding exercise to turn powers of A into vectors. Then use your CAS to find a spanning set (or basis, which is a special spanning set) for subspaces $V_k = \operatorname{span}\{A^0, A^1, \ldots, A^k\}, k = 1, 2, 3, 4, 5, 6$. Based on this evidence, how many matrices will be required for a span of V_k? (Remember that $A^0 = I$.)

3.3. Linear Combinations

We have seen in Section 3.2 that linear combinations give us a rich source of subspaces for a vector space. In this section we will take a closer look at properties of linear combinations.

Linear Dependence

First we need to make precise the idea of redundant vectors that we encountered in Example 3.2.10. About lists and sets: *Lists* involve an ordering of elements (they can just as well be called finite sequences), while *sets* don't really imply any ordering of elements. Thus, every list of vectors, e.g., $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, gives rise to a unique set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. A different list $\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_2$ may define the same set $\{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_2\} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Lists can have repeats in them, while sets don't. For instance, the list $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_1$ defines the set $\{\mathbf{v}_1, \mathbf{v}_2\}$. The terminology "the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$" really means "the set or a list of vectors consisting of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$"; the definitions below work perfectly well for either sets or lists.

DEFINITION 3.3.1. The vector $\mathbf{v}_i$ is *redundant* in the vectors $\mathbf{v}_1, \mathbf{v}_{2,}, \ldots, \mathbf{v}_n$ if the linear span $V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_{2,}, \ldots, \mathbf{v}_n\}$ does not change when $\mathbf{v}_i$ is removed.

EXAMPLE 3.3.2. Which vectors are redundant in the set consisting of $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$?

SOLUTION. As in Example 3.2.10, we notice that

$$\mathbf{v}_3 = (-1)\mathbf{v}_1 + 2\mathbf{v}_2$$

Thus any linear combination involving $\mathbf{v}_3$ can be expressed in terms of $\mathbf{v}_1$ and $\mathbf{v}_2$. Therefore, $\mathbf{v}_3$ is redundant in the list $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. But there is more going on here. Let's write the equation above in a form that doesn't single out any one vector:

$$0 = (-1)\mathbf{v}_1 + 2\mathbf{v}_2 + (-1)\mathbf{v}_3$$

Now we see that we could solve for *any* of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ in terms of the remaining two vectors. Therefore, each of these vectors is redundant in the set. However, this doesn't mean that we can discard all three and get the same linear span. This is obviously false. What we can do is discard any *one* of them, then start over and examine the remaining set for redundant vectors. □

This example shows that what really counts for redundancy is that the vector in question occur with a nonzero coefficient in a linear combination that equals 0. This situation warrants its own name:

Linearly Dependant or Independent Vectors

DEFINITION 3.3.3. The vectors $v_1, v_2, \ldots, v_n$ are said to be *linearly dependent* if there exist scalars $c_1, c_{2,}, \ldots, c_n$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = 0 \tag{3.3.1}$$

If these vectors are not linearly dependent, i.e., no nontrivial linear combination of them is equal to zero, then they are called *linearly independent*.

For convenience, we will call a linear combination *trivial* if it equals 0. Just as with redundancy, linear dependence or independence is a property of the list or set in question, *not* of the individual vectors. We are going to connect the ideas of linear dependence and redundancy. Here is the key fact.

Redundancy Test

THEOREM 3.3.4. *The list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of a vector space has redundant vectors if and only if it is linearly dependent, in which case the redundant vectors are those that occur with nonzero coefficient in some linear combination that sums to zero.*

PROOF. Observe that if (3.3.1) holds and some scalar, say c_1, is nonzero, then we can use the equation to solve for $\mathbf{v}_1$ as a linear combination of the remaining vectors:

$$\mathbf{v}_1 = \frac{-1}{c_1}(c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + \cdots + c_n\mathbf{v}_n)$$

Thus we see that any linear combination involving $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ can be expressed using only $\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$. It follows that

$$\operatorname{span}\{\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n\} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

Conversely, if these spans are equal then $\mathbf{v}_1$ belongs to the left hand side, so there are scalars $d_2, d_3, \ldots, d_n$ such that

$$\mathbf{v}_1 = d_2\mathbf{v}_2 + d_3\mathbf{v}_3 + \cdots + d_n\mathbf{v}_n$$

Now bring all terms to the right hand side and obtain the nontrivial linear combination

$$-\mathbf{v}_1 + d_2\mathbf{v}_2 + d_3\mathbf{v}_3 + \cdots + d_n\mathbf{v}_n = 0$$

All of this works equally well for any index other than 1 so the theorem is proved. □

It is instructive to examine the simple case of two vectors $\mathbf{v}_1, \mathbf{v}_2$. What does it mean to say that these vectors are linearly dependent? Simply that one of the vectors can be expressed in terms of the other. In other words, that each vector is a scalar multiple of the other. However, matters are more complex when we proceed to three or more vectors, a point that is often overlooked. So we issue a warning here.

Caution: If we know that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is linearly dependent, it does *not* necessarily imply that one of these vectors is a multiple of one of the others unless $n = 2$. In general, all we can say is that one of these vectors is a linear combination of the others.

EXAMPLE 3.3.5. Which of the following lists of vectors have redundant vectors, i.e., are linearly dependent?

$$\text{(a)} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

SOLUTION. Consider the vectors in each list to be designated as $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. Let's try to see the big picture. The general linear combination can be written as

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = A\mathbf{c}$$

where $A = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$ and $\mathbf{c} = (c_1, c_2, c_3)$. Now we see that a nontrivial linear combination that adds up to 0 amounts to a nontrivial solution to the homogeneous equation $A\mathbf{c} = 0$. We know how to find these! In case (a) we have that

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 0 & 1 & -2 \end{bmatrix} \xrightarrow[E_{21}(-1)]{} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 1 & -2 \end{bmatrix} \xrightarrow[E_{32}(-1)]{} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix}$$

so that the solutions to the homogeneous system are $c = (-c_3, 2c_3, c_3) = c_3(-1, 2, 1)$. Take $c_3 = 1$ and we have that

$$-1\mathbf{v}_1 + 2\mathbf{v}_2 + 1\mathbf{v}_3 = 0$$

which shows that $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is a linearly dependent list of vectors.

We'll solve (b) by a different method. Notice that

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

It follows that A is nonsingular, so the only solution to the system $A\mathbf{c} = 0$ is $c = 0$. Since every linear combination of the columns of A takes the form $A\mathbf{c}$, the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ must be linearly independent.

Finally, we see by inspection in (c) that since $\mathbf{v}_3$ is a repeat of $\mathbf{v}_1$, the linear combination

$$\mathbf{v}_1 + 0\mathbf{v}_2 - \mathbf{v}_3 = 0.$$

Thus, this list of vectors is linearly dependent. Notice, by the way, that not every coefficient c_i has to be nonzero. □

EXAMPLE 3.3.6. Show that any list of vectors that contains the zero vector is linearly dependent.

SOLUTION. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be such a list and suppose that for some index j, $\mathbf{v}_j = 0$. Examine the following linear combination which clearly sums to 0:

$$0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 1\mathbf{v}_j + \cdots 0\mathbf{v}_n = 0.$$

This linear combination is nontrivial because the coefficient of the vector $\mathbf{v}_j$ is 1. Therefore this list is linearly dependent by definition of dependence. □

The Basis Idea

We are now ready for one of the big ideas of vector space theory, the notion of a basis. We already know what a spanning set for a vector space V is. This is a set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ such that $V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. However, we saw back in Example 3.2.10 that some spanning sets are better than others because they are more economical. We know that a set of vectors has no redundant vectors in it if and only if it is linearly independent. This observation is the inspiration for the following definition.

DEFINITION 3.3.7. A *basis* for the vector space V is a spanning set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ which is a linearly independent set.

We should take note here that we could have just as well defined a basis as a minimal spanning set, by which we mean a spanning set such that any proper subset is not spanning set. The proof that this is equivalent to our definition of basis is left as an exercise.

Usually we think of a basis as a set of vectors and the order in which we list them is convenient but not important. Occasionally, ordering is important. In such a situation we speak of an *ordered basis* of $\mathbf{v}$ by which we mean a spanning list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ which is a linearly independent list.

EXAMPLE 3.3.8. Which subsets of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}\}$ yield bases of the vector space $\mathbb{R}^2$?

SOLUTION. These are just the vectors of Example 3.2.10 and Example 3.3.2. Referring back to that Example, we saw that

$$-\mathbf{v}_1 + 2\mathbf{v}_2 - \mathbf{v}_3 = 0$$

which told us that we could remove any one of these vectors and get the same span. Moreover, we saw that these three vectors span $\mathbb{R}^2$, so the same is true of any two of them. Clearly, a single vector cannot span $\mathbb{R}^2$ since the span of a single vector is a line through the origin. Therefore, the subsets $\{\mathbf{v}_1, \mathbf{v}_2\}$, $\{\mathbf{v}_2, \mathbf{v}_3\}$ and $\{\mathbf{v}_1, \mathbf{v}_3\}$ are all bases of $\mathbb{R}^2$. □

An extremely important generic type of basis is provided by the columns of the identity matrix. For future reference, we establish this notation.

NOTATION 3.3.9. The *standard basis* of $\mathbb{R}^n$ or $\mathbb{C}^n$ is the set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ where e_j is the column vector of size n whose jth entry is 1 and all other entries 0.

EXAMPLE 3.3.10. Let V be the standard vector space $\mathbb{R}^n$ or $\mathbb{C}^n$. Verify that the standard basis really is a basis of this vector space.

SOLUTION. Let $\mathbf{v} = (c_1, c_2, \ldots, c_n)$ be a vector from V so that $c_1, c_2, \ldots, c_n$ are scalars of the appropriate type. Now we have

$$\begin{aligned} \mathbf{v} &= \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + c_n \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \\ &= s_1\mathbf{e}_1 + s_2\mathbf{e}_2 + \cdots + s_n\mathbf{e}_n \end{aligned}$$

This equation tells us two things: first, every vector in V is a linear combination of the $\mathbf{e}_j$'s, so $V = \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$. Secondly, if some linear combination of vectors sums to the zero vector, then each scalar coefficient of the combination is 0. Therefore, these vectors are linearly independent. Therefore the set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a basis of V. □

Coordinates

In the case of the standard basis $\mathbf{e}_1, \mathbf{e}_2, , \mathbf{e}_3$ of $\mathbb{R}^3$ we know that it is very easy to write out any other vector $\mathbf{v} = (c_1, c_2, c_3)$ in terms of the standard basis:

$$\mathbf{v} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3$$

We call the scalars c_1, c_2, c_3 the *coordinates* of the vector $\mathbf{v}$. Up to this point, this is the only sense in which we have used the term "coordinates." We can see that these coordinates are strongly tied to the standard basis. Yet $\mathbb{R}^3$ has many bases. Is there a corresponding notion of "coordinates" relative to other bases? The answer is a definite "yes," thanks to the following fact.

THEOREM 3.3.11. *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be a basis of the vector space V. Then every $\mathbf{v} \in V$ can be expressed uniquely as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ up to order of terms.*

PROOF. To see this, note first that since

$$V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

there exist scalars $c_1, c_2, \ldots, c_n$ such that

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

Suppose also that we could write

$$\mathbf{v} = d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_n\mathbf{v}_n$$

Subtract these two equations and obtain

$$0 = (c_1 - d_1)\mathbf{v}_1 + (c_2 - d_2)\mathbf{v}_2 + \cdots + (c_n - d_n)\mathbf{v}_n$$

However, a basis is a linearly independent set, so it follows that each coefficient of this equation is zero, whence $c_j = d_j$, for $j = 1, 2, \ldots, n$. □

In view of this fact, we may speak of *coordinates of a vector relative to a basis*. Here is the notation that we employ:

DEFINITION 3.3.12. If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis of the vector space V and $\mathbf{v} \in V$ with $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$, then the scalars $c_1, c_2, \ldots, c_n$ are called the *coordinates of* $\mathbf{v}$ *with respect to the basis* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

As we have noted, coordinates of a vector with respect to the standard basis are what we have referred to as "coordinates" so far in this text. Perhaps we should call these the *standard coordinates* of a vector, but we will usually stick with the convention that an unqualified reference to a vectors coordinates assumes we mean standard coordinates. Normally vectors in $\mathbb{R}^n$ are given explicitly in terms of their standard coordinates, so these are trivial to identify. Coordinates with respect to other bases are fairly easy to calculate if we have enough information about the structure of the vector space.

EXAMPLE 3.3.13. The following vectors form a basis of $\mathbb{R}^3$: $\mathbf{v}_1 = (1, 1, 0)$, $\mathbf{v}_2 = (0, 2, 2)$ and $\mathbf{v}_3 = (1, 0, 1)$. Find the coordinates of $\mathbf{v} = (2, 1, 5)$ with respect to this basis.

SOLUTION. Notice that the basis $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ was given in terms of standard coordinates. Begin by writing

$$\begin{aligned}
\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 \\
&= [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}
\end{aligned}$$

where the coordinates c_1, c_2, c_3 of $\mathbf{v}$ relative to the basis $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are to be determined. This is a straightforward system of equations with coefficient matrix $A = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$ and right hand side $\mathbf{v}$. It follows that the solution we want is given by

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 0 \\ 0 & 2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}$$

This shows us that

$$\mathbf{v} = -1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

It does *not* prove that $\mathbf{v} = (-1, 1, 3)$, which is plainly false. Only in the case of the standard basis can we expect that a vector actually equals its vector of coordinates with respect to the basis. $\square$

In general, vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ are linearly independent if and only if the system $A\mathbf{c} = 0$ has only the trivial solution, where $A = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$. This in turn is equivalent to the matrix A being of full column rank n (see Theorem 2.5.9 of Chapter 2 where we see that these are equivalent conditions for a matrix to be invertible). We can see how this idea can be extended, and doing so tells us something remarkable. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be a basis of $V = \mathbb{R}^n$ and form the $n \times k$ matrix $A = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$. By the same reasoning as in the example, for any $\mathbf{b} \in V$ there is a unique solution to the system $A\mathbf{x} = \mathbf{b}$. In view of the Rank Theorem of Chapter 1 we see that A has full column rank. Therefore, $k \leq n$. On the other hand, we can take $\mathbf{b}$ to be any one of the standard basis vectors e_j, $j = 1, 2, \ldots, n$, solve the resulting systems and stack the solutions vectors together to obtain a solution to the system $AX = I_n$. From our rank inequalities, we see that

$$n = \operatorname{rank} I_n = \operatorname{rank} AX \leq \operatorname{rank} A = n$$

What this shows is that $k = n$, that is, every basis of $\mathbb{R}^n$ has exactly n elements in it. Does this idea extend to abstract vector spaces? Indeed it does, and we shall return to this issue in Section 3.5.

We are going to visit a problem which comes to us straight from calculus and analytical geometry (classification of conics) and show how the matrix and coordinate tools we have developed can shed light on this problem.

EXAMPLE 3.3.14. Suppose we want to understand the character of the graph of the curve $x^2 - xy + y^2 - 6 = 0$. It is suggested to us that if we that we execute a change of variables by rotating the xy-axis by $\pi/4$ to get a new $x'y'$-axis, the graph will become more intelligible. OK, we do it. The algebraic connection between the variables x, y and x', y', resulting from a rotation of θ can be worked out using a bit of trigonometry (which we omit) to yield

$$\begin{aligned} x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta \end{aligned}$$

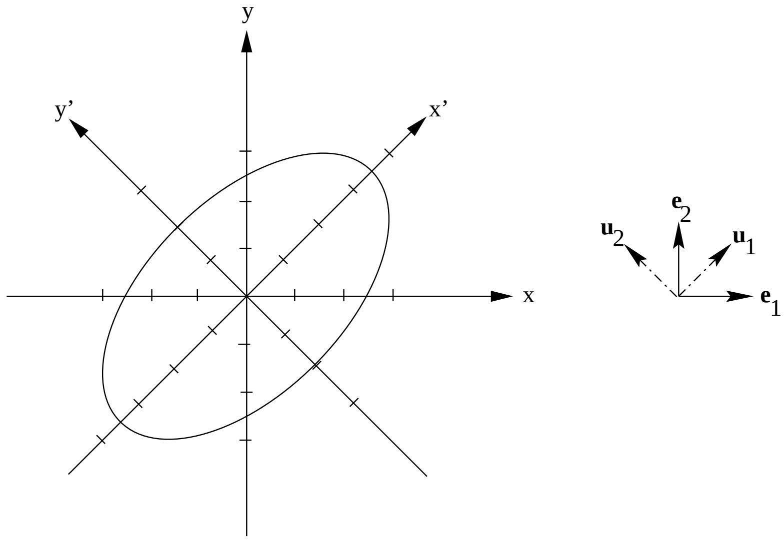


FIGURE 3.3.1. Change of variables and the curve $x^2 - xy + y^2 - 6 = 0$.

Use matrix methods to formulate these equations and execute the change of variables.

SOLUTION. First, we write the change of variable equations in matrix form as

$$\mathbf{x}' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = G(\theta)\mathbf{x}$$

We might recognize $G(\theta)$ as the Givens matrix introduced in Exericse x. This matrix isn't exactly what we need for substitution into our curve equation. Rather, we need x, y explicitly. That's easy enough. Simply invert $G(\theta)$ to obtain the rotation matrix $R(\theta)$ as

$$G(\theta)^{-1} = R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

so that $\mathbf{x} = R(\theta)\mathbf{x}'$. Now observe that the original equation can be put in the form

$$\begin{aligned} x^2 - xy + y^2 - 6 &= \mathbf{x}^T \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \mathbf{x} - 6 \\ &= \mathbf{x}^T \mathbf{R}(\theta)^T \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \mathbf{R}(\theta)\mathbf{x} - 6 \end{aligned}$$

We leave it as an exercise to check that with $\theta = \pi/4$ the equation reduces to $\frac{1}{2}(x'^2 + 3y'^2) - 6 = 0$ or equivalently

$$\frac{x'^2}{12} + \frac{y'^2}{4} = 1$$

This curve is simply an ellipse with axes $2\sqrt{3}$ and 2. With respect to the $x'y'$-axes, this ellipse is in so-called "standard form." For a graph of the ellipse, see Figure 3.3.1. $\square$

The change of variables we have just seen can be interpreted as a *change of coordinates* in the following sense. Notice that the variables x and y are just the standard coordinates (with respect to the standard basis $\mathbf{e}_1, \mathbf{e}_2$) of a general vector

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix} = x\mathbf{e}_1 + y\mathbf{e}_2$$

The meaning of the variables x' and y' becomes clear when we set $\mathbf{x}' = (x', y')$ and write the matrix equation $\mathbf{x} = R(\theta)\mathbf{x}'$ out in detail as a linear combination of the columns of $R(\theta)$:

$$\mathbf{x} = R(\theta)\mathbf{x}' = x' \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} + y' \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} = x'\mathbf{u}_1 + y'\mathbf{u}_2$$

Thus the numbers x' and y' are just the coordinates of the vector $\mathbf{x}$ with respect to a new basis $\mathbf{u}_1, \mathbf{u}_2$ of $\mathbb{R}^2$. This basis consists of unit vectors in the direction of the x' and y' axes. See Figure 3.3.1 for a picture of the two bases. Thus we see that the matrix $R(\theta)$ could reasonably be called a *change of coordinates* matrix or, as it is more commonly called, a *change of basis* matrix. Indeed, we can see from this example that the change of variables we encountered is nothing more than a change from one basis (the standard one, $\mathbf{e}_1, \mathbf{e}_2$) to another ($\mathbf{u}_1, \mathbf{u}_2$). The reason for a change of basis is that sometimes a problem looks a lot easier if we look at it using a basis other than the standard one, such as in our example. The concept of change of basis is explored in more generality in Section 3.7.

3.3 Exercises

1. Which of the following sets are linearly independent in $V = \mathbb{R}^3$? If not linearly independent, which vectors are redundant in the lists?

(a) $(1,0,1), (1,-1,1)$ (b) $(1,2,1), (2,1,1), (3,4,1), (2,0,1)$

(c) $(1,0,-1), (1,1,0), (1,-1,-2)$ (d) $(0,1,-1), (1,0,0), (-1,1,3)$

2. Which of the following sets are linearly independent in $V = \mathcal{P}_2$? If not linearly independent, which vectors are redundant in the lists?

(a) $x, 5x$ (b) $2, 2-x, x^2, 1+x^2$ (c) $1+x, 1+x^2, 1+x+x^2$ (d) $x-1, x^2-1, x+1$

3. Which of the following sets are linearly independent in $V = \mathbb{P}^3$? If not linearly independent, which vectors are redundant in the lists

(a) $1, x, x^2, x^3$ (b) $1+x, 1+x^2, 1+x^3$

(c) $1-x^2, 1+x, 1-x-2x^2$ (d) $x^2-x^3, x, -x+x^2+3x^3$

4. Find the coordinates of $\mathbf{v}$ with respect to the following bases:

(a) $\mathbf{v} = (0,1,2)$ with respect to basis $(2,0,1),\ (-1,1,0),\ (0,1,1)$ of $\mathbb{R}^3$.

(b) $\mathbf{v}$ $=2+x^2$ with respect to basis $1+x,\ x+x^2,\ 1-x$ of $\mathcal{P}_2$.

(c) $\mathbf{v} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ with respect to basis $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ of the space of real symmetric 2×2 matrices.

(d) $\mathbf{v} = (1,2)$ with respect to basis $(2+i,1),\ (-1,i)$ of $\mathbb{C}^2$.

5. In the following, $\mathbf{u}_1 = (1, 0, 1)$ and $\mathbf{u}_2 = (1, -1, 1)$.

(a) Determine if $\mathbf{v} = (2, 1, 2)$ belongs to the space $\mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

(b) Find a basis of $\mathbb{R}^3$ which contains $\mathbf{u}_1$ and $\mathbf{u}_2$.

6. In the following, $\mathbf{u}_1 = 1 - x + x^2$ and $\mathbf{u}_2 = x + 2x^2$.

(a) Determine if $\mathbf{v} = 4 - 7x - 2x^2$ belongs to the space $\mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

(b) Find a basis of $\mathbb{R}^3$ which contains $\mathbf{u}_1$ and $\mathbf{u}_2$.

7. If $2\mathbf{v}_1 + \mathbf{v}_3 + \mathbf{v}_4 = 0$ and $\mathbf{v}_2 + \mathbf{v}_3 = 0$ then what is the smallest spanning set $\mathrm{span}\,\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ may be reduced to?

8. Let $V = \mathbb{R}^{n,n}$ be the vector space of real $n \times n$ matrices and let $A, B \in \mathbb{R}^{n,n}$ such that both are nonzero matrices, A is nilpotent (some power of A is zero) and B is idempotent ($B^2 = B$). Show that the subspace $W = \mathrm{span}\{A, B\}$ cannot be spanned by a single element of W.

9. Let V be a vector space whose only subspaces are $\{\mathbf{0}\}$ and V. Show that V is the span of a single vector.

10. Suppose that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly independent elements of $\mathbb{R}^n$ and $A = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$. Show that $\mathrm{rank}\, A = k$.

11. Determine a largest subset of the following set of matrices which is linearly independent. Then expand this subset to a basis of $\mathbb{R}^{2,2}$.

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

12. The Wronskian of smooth functions $f(x), g(x), h(x)$ is defined as follows (a similar definition can be made for any number of functions):

$$W(f, g, h)(x) = \det \begin{bmatrix} f(x) & g(x) & h(x) \\ f'(x) & g'(x) & h'(x) \\ f''(x) & g''(x) & h''(x) \end{bmatrix}$$

Calculate the Wronskians of the three polynomial functions in parts (b) and (c) of Exercise 3.

13. Show that if $f(x), g(x), h(x)$ are linearly dependent smooth functions, then for all x, $W(f, g, h)(x) = 0$

14. Show that the functions e^x, x^3 and $\sin(x)$ are linearly independent in $C[0, 1]$ in two ways:

(a) Use Exercise 13.

(b) Assume a linear combination sums to zero and evaluate it at various points to obtain conditions on the coefficients.

15. Prove that a list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ with repeated vectors in it is linearly dependent.

16. Show that a linear operator $T : V \to W$ maps a linearly dependent set $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ to linearly dependent set $T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)$, but if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is linearly independent, $T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)$ need not be linearly independent (give a specific counterexample).

17. Let $R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ and $A = \begin{bmatrix} 1 & \frac{-1}{2} \\ \frac{-1}{2} & 1 \end{bmatrix}$. Calculate $R(\theta)^T A R(\theta)$ in the case that $\theta = \pi/4$.

18. Use matrix methods as in Example 3.3.14 to express the equation of the curve $11x^2 + 10\sqrt{3}xy + y^2 - 16 = 0$ in new variables x', y' obtained by rotating the xy-axis by an angle of $\pi/4$.

19. Suppose that a linear change of variables from old coordinates x_1, x_2 to new coordinates x_1', x_2' is defined by the equations

$$\begin{aligned} x_1 &= p_{11}x_1' + p_{12}x_2' \\ x_2 &= p_{21}x_1' + p_{22}x_2' \end{aligned}$$

where the 2×2 change of basis matrix $P = [p_{ij}]$ is invertible. Show that if a linear matrix multiplication function $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ is given in old coordinates by

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = T_A\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = T_A(\mathbf{x}) = A\mathbf{x}$$

where $A = [a_{ij}]$ is any 2×2 matrix, then it is given by $\mathbf{y}' = P^{-1}AP\mathbf{x}' = T_{P^{-1}AP}(\mathbf{x}')$ in new coordinates. *Hint:* Both domain and range elements $\mathbf{x}$ and $\mathbf{y}$ are given in terms of old coordinates. Express them in terms of new coordinates.

3.4. Subspaces Associated with Matrices and Operators

Certain subspaces are a rich source of information about the behavior of a matrix or a linear operator. We will define and explore the properties of these subspaces in this section.

Subspaces Defined by Matrices

Suppose we are given a matrix A. There are three very useful subspaces that can be associated with the matrix A. Understanding these subspaces is a great aid in vector space calculations that might have nothing to do with matrices per se, such as the determination of a minimal spanning set for a vector space. We shall follow each definition below by an illustration using the following example matrix.

$$A = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 0 & 1 & 2 & 1 \end{bmatrix} \tag{3.4.1}$$

We make the default assumption that the scalars are the real numbers, but the definitions we will give can be stated just as easily for the complex numbers.

DEFINITION 3.4.1. The *column space* of the $m \times n$ matrix A is the subspace $\mathcal{C}(A)$ of $\mathbb{R}^m$ spanned by the columns of A. Column Space

EXAMPLE 3.4.2. Describe the column space of the matrix A in Equation 3.4.1

SOLUTION. Here we have that $\mathcal{C}(A) \subseteq \mathbb{R}^2$. Also

$$\mathcal{C}(A) = \operatorname{span}\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}\}$$

Technically, this describes the column space in question, but we can do much better. We saw in Example 3.2.10 that the vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ was really redundant since it is a linear combination of the first two vectors. We also see that

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} = -2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

so that Theorem 3.2.9 shows us that

$$\mathcal{C}(A) = \operatorname{span}\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\}$$

This description is much better, in that it exhibits a *basis* of $\mathcal{C}(A)$. □

This example shows that not all the columns of the matrix A are really needed to span the entire subspace $\mathcal{C}(A)$. Clearly, this last expression for $\mathcal{C}(A)$ is much more economical than the first. How do you think it compares to the containing space $\mathbb{R}^2$?

DEFINITION 3.4.3. The *row space* of the $m \times n$ matrix A is the subspace $\mathcal{R}(A)$ of $\mathbb{R}^n$ spanned by the transposes of the rows of A.

The "transpose" part of the preceding definition seems a bit odd. Why would we want rows to look like columns? It's a technicality, but later it will be convenient for us to have the row spaces live inside a $\mathbb{R}^n$ instead of an $(\mathbb{R}^n)^T$. Remember, we had to make a choice about $\mathbb{R}^n$ consisting of rows or columns. Just to make the elements of a row space look like rows, we can always adhere to the tuple notation instead of matrix notation.

EXAMPLE 3.4.4. Describe the row space of A in Equation 3.4.1

SOLUTION. We have from definition that

$$\mathcal{R}(A) = \operatorname{span}\{(1, 1, 1, -1), (0, 1, 2, 1)\} \subseteq \mathbb{R}^4$$

Now it's easy to see that neither one of these vectors can be expressed as a multiple of the other (if we had $c(1, 1, 1, -1) = (0, 1, 2, 1)$, then read the first coordinates and obtain $c = 0$), so that span is given as economically as we can do, that is, the two vectors listed constitute a *basis* of $\mathcal{R}(A)$. □

DEFINITION 3.4.5. The *null space* of the $m \times n$ matrix A is the subset $\mathcal{N}(A)$ of $\mathbb{R}^n$ defined by

$$\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = 0\}$$

Observe that $\mathcal{N}(A)$ is the solution set to the homogeneous linear system $A\mathbf{x} = 0$. This means that null spaces are really very familiar. We were computing these solution sets back in Chapter 1. We didn't call them subspaces at the time. Here is an application of this concept. Let A be a square matrix. We know that A is invertible exactly when the system $A\mathbf{x} = 0$ has only the trivial solution (see Theorem 2.5.9). Now we can add one more equivalent condition to the long list of equivalences for invertibility: A is invertible exactly when $\mathcal{N}(A) = \{0\}$. Let us next justify the subspace property implied by the term "null space."

EXAMPLE 3.4.6. Use the subspace test to verify that $\mathcal{N}(A)$ really is a subspace of $\mathbb{R}^n$.

SOLUTION. Since $A0 = 0$, the zero vector is in $\mathcal{N}(A)$. Now let c be a scalar and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ arbitrary elements of $\mathcal{N}(A)$. By definition, $A\mathbf{u} = 0$ and $A\mathbf{v} = 0$. Add these two equations to obtain that

$$0 = 0 + 0 = A\mathbf{u} + A\mathbf{v} = A(\mathbf{u} + \mathbf{v})$$

Therefore $\mathbf{u} + \mathbf{v} \in \mathcal{N}(A)$. Next multiply the equation $A\mathbf{u} = 0$ by the scalar c to obtain

$$0 = c0 = c(A\mathbf{u}) = A(c\mathbf{u})$$

Thus we see from that definition that $c\mathbf{u} \in \mathcal{N}(A)$. The subspace test implies that $\mathcal{N}(A)$ is a subspace of $\mathbb{R}^n$. □

EXAMPLE 3.4.7. Describe the null space of the matrix A of Equation 3.4.1.

SOLUTION. Proceed as in Chapter 1. We find the reduced row echelon form of A, identify the free variables and solve for the bound variables using the implied zero right hand side and solution vector $x = [x_1, x_2, x_3, x_4]^T$:

$$\begin{bmatrix} 1 & 1 & 1 & -1 \\ 0 & 1 & 2 & 1 \end{bmatrix} \xrightarrow[E_{12}(-1)]{} \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 1 \end{bmatrix}$$

Pivots are in the first and second columns, so it follows that x_3 and x_4 are free, x_1 and x_2 are bound and

$$\begin{aligned} x_1 &= x_3 + 2x_4 \\ x_2 &= -2x_3 - x_4 \end{aligned}$$

Let's write out the form of a general solution, which will be strictly in terms of the free variables, and write the result as a combination of x_3 times some vector plus x_4 times another vector:

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

This is really a wonderful trick! Remember that free variables can take on arbitrary values, so we see that the general solution to the homogeneous system has the form of an arbitrary linear combination of the two vectors on the right. In other words,

$$\mathcal{N}(A) = \operatorname{span}\{ \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ 1 \end{bmatrix} \} \subseteq \mathbb{R}^4$$

Neither of these vectors is a multiple of the other, so this is as economical an expression for $\mathcal{N}(A)$ as we can hope for. In other words, we have exhibited a minimal spanning set, that is, a *basis* of $\mathcal{N}(A)$. □

EXAMPLE 3.4.8. Suppose that we have a sequence of vectors $\mathbf{x}^{(k)}$, $k = 0, 1, \ldots$. which are related by the formula

$$\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)}$$

where

$$\mathbf{x}^{(k)} = \begin{bmatrix} a_k \\ b_k \end{bmatrix} \qquad \text{and} \qquad A = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}$$

Further, suppose these vectors converge to some limiting vector x. What could the limiting vector be? (This example relates null spaces to the idea of a limiting state for a Markov chain as discussed in Example 2.3.4 of Chapter 2.)

SOLUTION. We reason as follows: since the limit of the sequence is $\mathbf{x}$, we can take the limits of both sides of the matrix equation above and obtain that

$$\mathbf{x} = A\mathbf{x} = I\mathbf{x}$$

so that

$$0 = \mathbf{x} - A\mathbf{x} = I\mathbf{x} - A\mathbf{x} = (I - A)\mathbf{x}$$

It follows that $\mathbf{x} \in \mathcal{N}(I - A)$. Now

$$I - A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.3 & -0.4 \\ -0.3 & 0.4 \end{bmatrix}$$

Calculate the null space by Gauss-Jordan elimination:

$$\begin{bmatrix} 0.3 & -0.4 \\ -0.3 & 0.4 \end{bmatrix} \xrightarrow[E_1(1/0.3)]{E_{21}(1)} \begin{bmatrix} 1 & -4/3 \\ 0 & 0 \end{bmatrix}$$

Therefore the null space of $I - A$ is spanned by the single vector $4/3, 1)$. In particular, any multiple of this vector qualifies as a possible limiting vector. If we want a limiting vector whose entries are nonnegative and sum to 1, then the only choice is the vector resulting from dividing $(4/3, 1)$ by the sum of its coordinates to obtain

$$(3/7)(4/3, 1) = (4/7, 3/7) \approx (0.57143, 0.42857).$$

Interestingly enough, this is the vector that resulted from the calculation on page 66. □

An important point that comes out of the previous examples is that we can express a null space as a simple linear span by using the methods for system solving we developed in Chapter 1, together with a little algebra.

We conclude this excursion on subspaces with an extremely powerful way of thinking about the product $A\mathbf{x}$, which was first introduced in Example 2.2.6 of Chapter 2. We shall use this idea often.

THEOREM 3.4.9. *Let the matrix A have columns* $\mathbf{a}_1, \ldots, \mathbf{a}_n$, *i.e.,* $A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$. *Let* $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$. *Then*

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

In other words, $A\mathbf{x}$ is simply a linear combination of the columns of A with vector of coefficients $\mathbf{x}$.

PROOF. As usual, let $A = [a_{ij}]$. The ith entry of the vector $\mathbf{v} = A\mathbf{x}$ is, by definition of matrix multiplication

$$v_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n$$

However, $\mathbf{a}_j = [a_{1j}, a_{2j}, \ldots, a_{nj}]^T$ so that the ith entry of the linear combination $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$ is

$$x_1a_{i1} + x_2a_{i2} + \cdots + x_na_{in} = v_i$$

which is what we wanted to show. □

This theorem shows that the column space of the matrix A can be thought of as the set of all possible matrix products Ax, i.e.,

$$\mathcal{C}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$$

Furthermore, the coefficients of a linear combination $A\mathbf{x}$ are precisely the entries of the vector $\mathbf{x}$. Two important insights follow from these observations: firstly, we see that the linear system $A\mathbf{x} = \mathbf{b}$ is consistent exactly when $b \in \mathcal{C}(A)$. Secondly, we see that the linear combination of columns of A with coefficients from the vector $\mathbf{x}$ is trivial, i.e., the combination sums to the zero vector, exactly when $\mathbf{x} \in \mathcal{N}(A)$.

EXAMPLE 3.4.10. Interpret the linear combination

$$\mathbf{v} = 3\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 5\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

as a matrix multiplication.

SOLUTION. By using Theorem 3.4.9 we can obtain several interpretations of this linear combination. The most obvious is

$$\mathbf{v} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$$

□

Notice that the linear combination of the previous example is not trivial, i.e., does not sum to the zero vector. Also, if we were working with the matrix of Equation 3.4.1, we could obtain a linear combination with the same value as the previous example in the form

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

This gives the same result.

Subspaces Defined by a Linear Operator

Suppose we are given a linear operator $T : V \to W$. We immediately have three spaces we can associate with the operator, namely, the domain V, target W and range $\operatorname{range}(T) = \{\mathbf{y} \mid \mathbf{y} = T(\mathbf{x}) \text{ for some } \mathbf{x} \in V\}$ of the operator. The domain and range are vector spaces by definition of linear operator. That that range is a vector space is a nice example of using the subspace test.

EXAMPLE 3.4.11. Show that if $T : V \to W$ is a linear operator, then $\operatorname{range}(T)$ is a subspace of W.

SOLUTION. Apply the subspace test. First, we observe that $\operatorname{range}(T)$ contains $T(0)$. We leave it as an exercise for the reader to check that $T(0)$ is the zero element of W. Next let $\mathbf{y}$ and $\mathbf{z}$ be in $\operatorname{range}(T)$, say $\mathbf{y} = T(\mathbf{u})$ and $\mathbf{z} = T(\mathbf{v})$. We show closure of $\operatorname{range}(T)$ under addition: by the linearity property of T

$$\mathbf{y} + \mathbf{z} = T(\mathbf{u}) + T(\mathbf{v}) = T(\mathbf{u} + \mathbf{v}) \in \operatorname{range}(T)$$

where the latter term belongs to $\operatorname{range}(T)$ by definition of image. Finally, we show closure under scalar multiplication: let c be a scalar and we obtain from the linearity property of T that

$$c\mathbf{y} = cT(\mathbf{u}) = T(c\mathbf{u}) \in \operatorname{range}(T)$$

where the latter term belongs to $\operatorname{range}(T)$ by definition of range. Thus, the subspace test shows that $\operatorname{range}(T)$ is a subspace of W. □

Here is another space that has proven to be very useful in understanding the nature of a linear operator.

DEFINITION 3.4.12. The *kernel* of the linear operator $T : V \to W$ is the subspace of V given by

$$\ker(T) = \{\mathbf{x} \in V \mid T(\mathbf{x}) = 0\}$$

The definition claims that the kernel is a subspace and not merely a subset of the domain. The proof of this is left to the exercises. The fact is that we have been computing kernels since the beginning of the text. To see this, suppose that we have a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ given by matrix multiplication, that is, $T_A(\mathbf{x}) = A\mathbf{x}$, for all $x \in \mathbb{R}^n$. Then

$$\begin{aligned}\ker(T) &= \{\mathbf{x} \in \mathbb{R}^n \mid T_A(\mathbf{x}) = 0\} \\ &= \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = 0\} \\ &= \mathcal{N}(A)\end{aligned}$$

In other words, for matrix multiplications kernels are the same thing as null spaces.

Here is one really nice application of kernels. Suppose we are interested in knowing whether or not a given operator $T : V \to W$ is one-to-one, i.e., if the equation $T(\mathbf{u}) = T(\mathbf{v})$ necessarily implies that $\mathbf{u} = \mathbf{v}$. In general, this is a nontrivial question. If, for example, $V = W = \mathbb{R}$, then we could graph the function T and try to determine if a horizontal line cut the graph twice. But for *linear* operators, the answer is much simpler:

THEOREM 3.4.13. *If $T : V \to W$ is a linear operator, then T is one-to-one if and only if $\ker(T) = \{0\}$.*

PROOF. If T is one-to-one, then only one element can map to 0 under T. Thus, $\ker(T)$ can consist of only one element. However, we know that $\ker(T)$ contains the zero vector since it is a subspace of the domain of T. Therefore, $\ker(T) = \{0\}$.

Conversely, suppose that $\ker(T) = \{0\}$. If $\mathbf{u}$ and $\mathbf{v}$ are such that $T(\mathbf{u}) = T(\mathbf{v})$, then subtract terms and use the linearity of T to obtain that

$$0 = T(\mathbf{u}) - T(\mathbf{v}) = T(\mathbf{u}) + (-1)T(\mathbf{v}) = T(\mathbf{u} - \mathbf{v}).$$

It follows that $\mathbf{u} - \mathbf{v} \in \ker(T) = \{0\}$. Therefore, $\mathbf{u} - \mathbf{v} = 0$ and hence $\mathbf{u} = \mathbf{v}$. □

Before we leave the topic of one-to-one linear mappings, let's digest its significance in a very concrete case. The space $\mathcal{P}_2 = \operatorname{span}\{1, x, x^2\}$of polynomials of degree at most 2 has a basis of three elements, like $\mathbb{R}^3$ and it seems very reasonable to think that $\mathcal{P}_2$ is "just like" $\mathbb{R}^3$ in that a polynomial $p(x) = a + bx + cx^2$ is uniquely described by its vector of coefficients $(a, b, c) \in \mathbb{R}^3$ and corresponding polynomials and vectors add and scalar multiply in a corresponding way. Here is the precise version of these musings: define an operator $T : \mathcal{P}_2 \to \mathbb{R}^3$ by the formula $T(a + bx + cx^2) = (a, b, c)$. One can check that T is linear, the range of T is its target, $\mathbb{R}^3$, and $\ker(T) = 0$. By Theorem 3.4.13 the function T is one-to-one and maps its domain onto its target. Hence, it describes a one-to-one correspondence between elements of $\mathcal{P}_2$ and elements of $\mathbb{R}^3$ such that sums and scalar products in one space correspond to the corresponding sums and scalar products in the other. In plain words, this means we can get one of the vector spaces from the other simply by relabelling elements of one of the spaces. So, in a very real sense, they are "the same thing." More generally, whenever there is a one-to-one linear mapping of one vector space onto another, we say the two vector spaces are *isomorphic*, which is a fancy way of saying that they are the same, up to a relabelling of elements. A one-to-one and onto linear mapping, like our T, is called an *isomorphism*.

In summary, there are four important subspaces associated with a linear operator, the domain, target, kernel and range. In symbols

$$\begin{aligned}
\operatorname{domain}(T) &= V \\
\operatorname{target}(T) &= W \\
\ker(T) &= \{\mathbf{v} \in V \,|\, T(\mathbf{v}) = 0\} \\
\operatorname{range}(T) &= \{T(\mathbf{v}) \,|\, \mathbf{v} \in V\}
\end{aligned}$$

There are important connections between these subspaces and those associated with a matrix. Let A be an $m \times n$ matrix and $T_A : \mathbb{R}^n \to \mathbb{R}^m$ the corresponding operator defined by matrix multiplication by A. Then

$$\begin{aligned}
\operatorname{domain}(T_A) &= \mathbb{R}^n \\
\operatorname{target}(T_A) &= \mathbb{R}^m \\
\ker(T_A) &= \mathcal{N}(A) \\
\operatorname{range}(T_A) &= \mathcal{C}(A)
\end{aligned}$$

The proofs of these are left to the exercises. One last example of subspaces associated with a linear operator $T : V \to W$ is really a whole family of subspaces. Suppose that U is a subspace of the domain V. Then we define the *image of* U *under* T to be the set

$$T(U) = \{T(u) \mid u \in U\}$$

Interestingly enough, $T(U)$ is always a subspace of $\mathrm{range}(T)$. We leave the proof of this fact as an exercise. In words, what it says is that a linear operator maps subspaces of its domain into subspaces of its range.

3.4 Exercises

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

(a) Find the reduced row echelon form of the matrix A .

(b) Find a spanning set for the null space of A. *Hint:* See Example 3.4.7.

(c) Find a spanning set for the column space of A.

(d) Find a spanning set for the row space of A.

(e) Find all possible linear combinations of the columns of A that add up to $\mathbf{0}$. *Hint:* See the remarks following Theorem 3.4.9.

2. Let $A = \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 1 & 2 & 1 & 1 & 1 \\ 3 & 6 & 2 & 2 & 3 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 7 \\ \alpha \\ \beta \end{bmatrix}$ define the system $A\mathbf{x} = \mathbf{b}$. First find the reduced row echelon form for the augmented matrix $[A|\mathbf{b}]$. Then answer questions (a)-(e) of Exercise 1 and also

(f) For what values of α and β is the vector $\mathbf{b}$ in the column space of A? *Hint:* The remarks following Theorem 3.4.9 are helpful.

3. Repeat Exercise 1 with $A = \begin{bmatrix} 1 & i & 0 \\ 1 & 2 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$.

4. Let $\mathbf{u} = [1, 1, 0]^T$, $\mathbf{v} = [0, 1, 1]^T$, and $\mathbf{w} = [1, 3 - 2i, 1]^T$. Express the expression $2\mathbf{u} - 4\mathbf{v} - 3\mathbf{w}$ as a single matrix product.

5. The matrix $\begin{bmatrix} 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix}$ is a transition matrix for a Markov chain. Find the null space of $I - A$ and determine all state vectors (nonnegative entries that sum to 1) in the null space.

6. Show the range of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 - 2x_2 + x_3 \\ x_1 + x_2 + x_3 \\ 2x_1 - x_2 + 2x_3 \end{bmatrix}$$

is not $\mathbb{R}^3$, and find a vector not in the range. Is T one-to-one? Give reasons.

7. Show that if $T : V \to W$ is a linear operator, then $T(\mathbf{0}) = \mathbf{0}$.

8. Show that if $T : V \to W$ is a linear operator, then the kernel of T is a subspace of V.

9. Prove that if T is a linear operator, then for all $\mathbf{u}, \mathbf{v}$ in the domain of T and scalars c and d, we have $T(c\mathbf{u} - d\mathbf{v}) = cT(\mathbf{u}) - dT(\mathbf{v})$.

10. Prove that if A is a nilpotent $n \times n$ matrix then $\mathcal{N}(A) \neq \{\mathbf{0}\}$ and $\mathcal{N}(I - A) = \{\mathbf{0}\}$.

11. Let the function $T : \mathbb{R}^3 \to \mathcal{P}_2$ be defined by

$$T(\begin{bmatrix} a \\ b \\ c \end{bmatrix}) = ax + b(x-1) + cx^2$$

(a) Show that T is a linear operator.

(b) Show that $\ker T = \{\mathbf{0}\}$.

(c) Show that $\operatorname{range} T = \mathcal{P}_2$.

(d) Conclude that $\mathbb{R}^3$ and $\mathcal{P}_2$ are isomorphic vector spaces.

12. Let $T : V \to W$ be a linear operator and U a subspace of V. Show that the image of $U, T(U) = \{T(\mathbf{v}) \mid \mathbf{v} \in U\}$, is a subspace of W.

13. Let $T_A : \mathbb{R}^n \to \mathbb{R}^m$ be the matrix multiplication operator defined by the $m \times n$ matrix A. Show that $\ker T_A = \mathcal{N}(A)$ and $\operatorname{range} T = \mathcal{C}(A)$.

3.5. Bases and Dimension

We have used the word “dimension” many times already, without really making the word precise. Intuitively, it makes sense when we say that $\mathbb{R}^2$ is “two dimensional” or that $\mathbb{R}^3$ is “three dimensional”: we reason that it takes two coordinate numbers to determine a vector in $\mathbb{R}^2$ and three for a vector in $\mathbb{R}^3$. What can we say about general vector spaces? Is there some number that is a measure of the size of the vector space? These are questions we propose to answer in this section. In the familiar cases of geometrical vector spaces, the answers will merely confirm our intuition. The answers will also enable us to solve this kind of problem, which is a bit more subtle: suppose we solve a linear system of 5 equations in 10 unknowns and obtain through Gauss-Jordan elimination a solution that involves 4 free variables, which means that we can express all 10 unknowns in terms of these 4 free variables. Is it somehow possible, perhaps by using a totally different method of system solving, to arrive at a similar kind of solution that involves fewer free variables, say 3? This would, in effect, reduce the “degrees of freedom” of the system.

We aren’t going to answer this question just yet. This example is rather vague; nonetheless, our intuition might suggest that if the free variables really are independent, we shouldn't be able to reduce their number. Therefore the answer to the question of this example should be "no." What we are really asking is a question about the nature of the solution set of the system. As a matter of fact, our intuition is correct. Reasons for this answer will be developed in this section.

The Basis Theorem

We now know that the standard vector spaces always have a basis. Given an abstract vector space V, can we be sure that a basis for V exists? For the type of vector spaces that we introduced in Section 1, that is, subspaces of the standard vector spaces, we will see that the answer is an unconditional "yes." For the more general concept of an abstract vector space the answer is "sometimes." The following concept turns out to be helpful.

Finite Dimensional Vector Space

DEFINITION 3.5.1. The vector space V is called *finite dimensional* if there is a finite set of vectors $\{v_1, v_2, \ldots, v_n\}$ which is a spanning set for V.

Examples of finite dimensional vector spaces are the standard spaces $\mathbb{R}^n$ or $\mathbb{C}^n$. As a matter of fact, we will see shortly that every subspace of a finite dimensional vector space is finite dimensional, which includes most of the vector spaces we have studied so far. However, some very important vector spaces are *not* finite dimensional, and accordingly, we call them *infinite dimensional* spaces. Here is an example.

EXAMPLE 3.5.2. Show that the space of all polynomial functions $\mathcal{P}$ is not a finite dimensional space, while the subspaces $\mathcal{P}_n$ are finite dimensional.

SOLUTION. If $\mathcal{P}$ were a finite dimensional space, then there would be a finite spanning set of polynomials $p_1(x), p_2(x), \ldots, p_m(x)$ for $\mathcal{P}$. This means that any other polynomial could be expressed as a linear combination of these polynomials. Let m be the maximum of all the degrees of the polynomials $p_j(x)$. Notice that any linear combination of polynomials of degree at most m must itself be a polynomial of degree at most m. (Remember that polynomial multiplication plays no part here, only addition and scalar multiplication.) Therefore, it is not possible to express the polynomial $q(x) = x^{m+1}$ as a linear combination of these polynomials, which means that they cannot be a basis. Hence, the space $\mathcal{P}$ has no finite spanning set.

On the other hand, it is obvious that the polynomial

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n$$

is a linear combination of the monomials $1, x, \ldots, x^n$ from which it follows that $\mathcal{P}_n$ is a finite dimensional space. $\square$

Here is the first basic result about these spaces. It is simply a formalization of what we have already done with preceding examples.

Basis Theorem

THEOREM 3.5.3. *Every finite dimensional vector space has a basis.*

PROOF. To see this, suppose that V is a finite dimensional vector space with

$$V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

Now if the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ has a redundant vector in it, discard it and obtain a smaller spanning set of V. Continue discarding vectors until you reach a spanning set for V that has no redundant vectors in it. (Since you start with a finite set, this can't go on indefinitely.) By the redundancy test, this spanning set must be linearly independent. Hence it is a basis of V. □

The Dimension Theorem

No doubt you have already noticed that every basis of the vector space $\mathbb{R}^2$ must have exactly two elements in it. Similarly, one can reason geometrically that any basis of $\mathbb{R}^3$ must consist of exactly three elements. These numbers somehow measure the "size" of the space in terms of the degrees of freedom (number of coordinates) one needs to describe a general vector in the space. The content of the dimension theorem is that this number can be unambiguously defined. First, we need a very handy theorem which is sometimes called the Steinitz substitution principle. This principle is a mouthful to swallow, so we will precede its statement with an example that illustrates the basic idea.

EXAMPLE 3.5.4. Let $\mathbf{w}_1 = (1, -1, 0), \mathbf{w}_2 = (0, -1, 1), \mathbf{v}_1 = (0, 1, 0), \mathbf{v}_2 = (1, 1, 0)$, $\mathbf{v}_3 = (0, 1, 1)$. Then $\mathbf{w}_1, \mathbf{w}_2$ form a linearly independent set and $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ form a basis of $V = \mathbb{R}^3$ (assume this). Show how to substitute both $\mathbf{w}_1$ and $\mathbf{w}_2$ into the set $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ while substituting out some of the $\mathbf{v}_j$ and at the same time retaining the basis property of the set.

SOLUTION. Since $\mathbb{R}^3 = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, we can express $\mathbf{w}_1$ as a linear combination of these vectors. We have a formal procedure for finding such combinations, but in this case we don't have to work too hard. A little trial and error shows

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = -1 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = -1\mathbf{v}_1 + 2\mathbf{v}_2 + 0\mathbf{v}_3$$

so that $1\mathbf{w}_1 + 1\mathbf{v}_1 - 2\mathbf{v}_2 - 0\mathbf{v}_3 = 0$. It follows that $\mathbf{v}_1$ or $\mathbf{v}_2$ is redundant in the set $\mathbf{w}_1, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. So discard, say, $\mathbf{v}_2$, and obtain a spanning set $\mathbf{w}_1, \mathbf{v}_1, \mathbf{v}_3$. In fact, it is actually a basis of V since two vectors can only span a plane. Now start over: express $\mathbf{w}_2$ as a linear combination of this new basis. Again, a little trial and error shows

$$\mathbf{w}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} = -2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = 0\mathbf{w}_1 - 2\mathbf{v}_1 + 1\mathbf{v}_3$$

Therefore $\mathbf{v}_1$ or $\mathbf{v}_3$ is redundant in the set $\mathbf{w}_1, \mathbf{w}_2, \mathbf{v}_1, \mathbf{v}_3$. So discard, say, $\mathbf{v}_3$, and obtain a spanning set $\mathbf{w}_1, \mathbf{w}_2, \mathbf{v}_1$. Again, this set is actually a basis of V since two vectors can only span a plane; and this is the kind of set we were looking for. □

Steinitz Substitution Principle

THEOREM 3.5.5. *Let $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ be a linearly independent set in the space V and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be a basis of V. Then $r \leq n$ and we may substitute all of the $\mathbf{w}_i$'s for some of the $\mathbf{v}_j$'s in such a way that the resulting set of vectors is still a basis of V.*

PROOF. Let's do the substituting one step at a time. Suppose that $k < r$ and that we have relabelled the remaining $\mathbf{v}_i$'s so that

$$V = \operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \dots \mathbf{w}_k, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s\}$$

with $k + s \leq n$ and $\mathbf{w}_1, \mathbf{w}_2, \dots \mathbf{w}_k, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s$ a basis of V. (Notice that $k = 0$ and $s = n$ when we start, so $k + s = n$.)

We show how to substitute the next vector $\mathbf{w}_{k+1}$ into the basis. Certainly

$$V = \operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \dots \mathbf{w}_k, \mathbf{w}_{k+1}, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s\}$$

as well, but this spanning set is linearly dependent since $\mathbf{w}_{k+1}$ is linearly dependent on $\mathbf{w}_1, \mathbf{w}_2, \dots \mathbf{w}_k, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s$. Also, there have to be some $\mathbf{v}_i$s left if $k < r$, for otherwise a proper subset of the $\mathbf{w}_j$s would be a basis of V. Now use the redundancy test to discard, one at a time, as many of the $\mathbf{v}_j$'s from this spanning set as possible, all the while preserving the span. Again relabel the $\mathbf{v}_j$'s that are left so as to obtain for some $t \leq s$ a spanning set

$$\mathbf{w}_1, \mathbf{w}_2, \dots \mathbf{w}_{k+1}, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_t$$

of V from which no $\mathbf{v}_j$ can be discarded without shrinking the span. Could this set be linearly dependent? If so, there must be some equation of linear dependence among the vectors such that none of the vectors $\mathbf{v}_j$ occurs with a nonzero coefficient; otherwise, according to the redundancy test, such a $\mathbf{v}_j$ could be discarded and the span preserved. Therefore, there is an equation of dependency involving only the $\mathbf{w}_j$'s. This means that the vectors $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r$ form a linearly dependent set, contrary to hypothesis. Hence, there is no such linear combination and the vectors $\mathbf{w}_1, \mathbf{w}_2, \dots \mathbf{w}_{k+1}, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_t$ are linearly independent, as well as a spanning set of V. Now we have to have discarded at least one of the $\mathbf{v}_i$'s since $\mathbf{w}_1, \mathbf{w}_2, \dots \mathbf{w}_k, \mathbf{w}_{k+1}, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s$ is a linearly dependent set. Therefore, $t \leq s - 1$. It follows that

$$\begin{aligned}(k+1) + t &\leq k + 1 + s - 1\\ &\leq k + s\\ &\leq n\end{aligned}$$

Now continue this process until $k = r$. □

Dimension Theorem

THEOREM 3.5.6. *Let V be a finite dimensional vector space. Then any two bases of V have the same number of elements which is called the dimension of the vector space and denoted as* $\dim V$.

PROOF. Let $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r$ and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ be two given bases of V. Apply the Steinitz substitution principle to the linearly independent set $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r$ and the basis $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ to obtain that $r \leq n$. Now reverse the roles of these two sets in the substitution principle to obtain the reverse inequality $n \leq r$. We conclude that $r = n$, as desired. □

Remember that a vector space always carries a field of scalars with it. If we are concerned about that field we could specify it explicitly as part of the dimension notation. For instance, we could write

$$\dim \mathbb{R}^n = \dim_{\mathbb{R}} \mathbb{R}^n \text{ or } \dim \mathbb{C}^n = \dim_{\mathbb{C}} \mathbb{C}^n.$$

Usually, the field of scalars is clear and we don't need the subscript notation.

As a first application, let's dispose of the standard spaces. We already know from Example 3.3.10 that these vector spaces have a basis consisting of n elements, namely the standard basis $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. According to the dimension theorem, this is all we need to specify dimension of these spaces.

COROLLARY 3.5.7. *For the standard spaces we have*

$$\begin{aligned} \dim \mathbb{R}^n &= n \\ \dim \mathbb{C}^n &= n \end{aligned}$$

There is one more question we want to answer right away. How do dimensions of a finite dimensional vector space V and a subspace W of V relate to each other? Actually, we don't even know if W is finite dimensional. Our intuition tells us that subspaces should have smaller dimension. Sure enough, our intuition is right this time! The tool that we use to confirm this fact is useful in its own right.

COROLLARY 3.5.8. *If W is a subspace of the finite dimensional vector space V, then W is also finite dimensional and*

$$\dim W \leq \dim V$$

with equality if and only if $V = W$.

PROOF. Let $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ be a linearly independent set in W and suppose that $\dim V = n$. According to the Steinitz substitution principle, $r \leq n$. So there is an upper bound on the number of elements of a linearly independent set in W. Now if we had that the span of $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ is smaller than W, then we could find a vector $\mathbf{w}_{r+1}$ in W but not in $\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_r\}$. The new set $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r, \mathbf{w}_{r+1}$ would also be linearly independent (we leave this as an exercise). Since we cannot continue adding vectors indefinitely, we have to conclude that at some point we obtain a basis $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_s$for W. Furthermore, $s \leq n$, so we conclude that W is finite dimensional and $\dim W \leq \dim V$. Finally, if we had equality, then a basis of W would be the same size as a basis of V. However, Steinitz substitution ensures that any linearly independent set can be expanded to a basis of V. It follows that a basis for W is also a basis for V, whence $W = V$. □

3.5 Exercises

1. Find all possible subsets of the following sets of vectors that form a basis of $\mathbb{R}^3$.

(a) $(1, 0, 1), (1, -1, 1)$ (b) $(1, 2, 1), (2, 1, 1), (3, 4, 1), (2, 0, 1)$

(c) $(2, -3, 1), (4, -2, -3), (1, 1, 1)$

2. The sets of vectors listed below form bases and linearly independent sets in their respective spaces. According to Steinitz substitution , the $\mathbf{w}_1$'s may be substituted in for some $\mathbf{v}_j$'s and retain the basis property. Which $\mathbf{v}_j$'s could be replaced by $\mathbf{w}_1$and $\mathbf{w}_2$.

(a) In $\mathbb{R}^3$, $\mathbf{v}_1 = (1, 3, 1), \mathbf{v}_2 = (2, -1, 1), \mathbf{v}_3 = (1, 0, 1)$ and $\mathbf{w}_1 = (0, 1, 0), \mathbf{w}_2 = (1, 1, 1)$

(b) In $\mathcal{P}_2$, $\mathbf{v}_1 = 1 - x, \mathbf{v}_2 = 2 + x, \mathbf{v}_3 = 1 + x^2$ and $\mathbf{w}_1 = x, \mathbf{w}_2 = x^2$

3. If U and V are subspaces of the finite dimensional vector space W and $U \cap V = \{\mathbf{0}\}$, prove that $\dim(U + V) = \dim U + \dim V$.

4. Let $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ be linearly independent vectors in the vector space W. Show that if $\mathbf{w}$ is a vector in W and $\mathbf{w} \notin \operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$, then $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r, \mathbf{w}$ is a linearly independent set.

5. Answer True/False to each part. In what follows, assume that V is a vector space of dimension n and $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \subseteq V$.

(a) If S is a basis of V then $k = n$.

(b) If S spans V then $k \leq n$.

(c) If S is linearly independent then $k \leq n$.

(d) If S is linearly independent and $k = n$ then S spans V.

(e) If S spans V and $k = n$ then S is a basis for V.

(f) If A is a 5 by 5 matrix and $\det A = 2$, then the first 4 columns of A span a 4 dimensional subspace of $\mathbb{R}^5$.

(g) A linearly independent set contains redundant vectors.

(h) If $V = \operatorname{span}\{\mathbf{v}_2, \mathbf{v}_3\}$ and $\dim V = 2$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly dependent set.

(i) A set of vectors containing the zero vector is a linearly independent set.

(j) Every vector space is finite dimensional.

(k) The set of vectors $[i, 0]^T, [0, i]^T, [1, i]^T$ in $\mathbb{C}^2$ contains redundant vectors.

6. Show that the functions $1, x, x^2, \ldots, x^n$ form a basis for the space $\mathcal{P}_n$ of polynomials of degree at most n.

7. Prove that $C[0, 1]$ is an infinite dimensional space (*Hint:* $\mathcal{P}$ is a subspace of $C[0, 1]$).

8. Let $E_{i,j}$ be the $m \times n$ matrix with a unit in the (i, j)th entry and zeros elsewhere. Prove that $\{E_{i,j} \mid i = 1, \ldots, m,\ j = 1, \ldots, n\}$ is a basis of the vector space $\mathbb{R}^{m,n}$.

9. Let $T : V \to W$ be a linear operator such that $\operatorname{range} T = W$ and $\ker T = \{0\}$. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be a basis of V. Show that $T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)$ is a basis of W.

10. Let $V = \{\mathbf{0}\}$, a vector space with a single element. Explain why the element $\mathbf{0}$ is *not* a basis of V and the dimension of V must be 0.

11. Show that a set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in the vector space V is a basis if and only if it is a minimal spanning set, that is, no proper subset of these vectors is a spanning set.

12. Let $T : V \to W$ be a linear operator where V is a finite dimensional space and U is a subspace of V. Prove that $\dim T(U) \leq \dim U$.

13. Determine the dimension of the subspace of $\mathbb{R}^{n,n}$ consisting of all symmetric matrices by exhibiting a basis.

14. Let U be the subspace of $W = \mathbb{R}^{n,n}$ consisting of all symmetric matrices and V the subspace of all skew-symmetric matrices.

(a) Show that $U \cap V = \{\mathbf{0}\}$.

(b) Show that $U + V = W$.

(c) Use Exercises 8, 3 and 13 to calculate the dimension of V.

3.6. Linear Systems Revisited

We now have some very powerful tools for understanding the nature of solution sets of the standard linear system $A\mathbf{x} = \mathbf{b}$. This understanding will help us design practical computational methods for finding dimension and bases for vector spaces and other problems as well.

The first business at hand is to describe solution sets of non-homogeneous systems. Recall that every homogeneous system is consistent since it has the trivial solution. Nonhomogeneous systems are another matter. We already have one criterion, namely that rank of augmented matrix and coefficient matrix of the system must agree. Here is one more way to view the consistency of such a system in the language of vector spaces.

THEOREM 3.6.1. *The linear system $A\mathbf{x} = \mathbf{b}$ of m equations in n unknowns is consistent if and only if $\mathbf{b} \in \mathcal{C}(A)$.*

Consistency in Terms of Column Space

PROOF. The key to this fact is Theorem 3.4.9, which says that the vector $A\mathbf{x}$ is a linear combination of the columns of A with the entries of $\mathbf{x}$ as scalar coefficients. Therefore, to say that $A\mathbf{x} = \mathbf{b}$ has a solution is simply to say that some linear combination of columns of A adds up to $\mathbf{b}$, i.e., $\mathbf{b} \in \mathcal{C}(A)$. □

EXAMPLE 3.6.2. One of the following vectors belongs to the space V spanned by $\mathbf{v}_1 = (1, 1, 3, 3)$, $\mathbf{v}_2 = (0, 2, 2, 4)$ and $\mathbf{v}_3 = (1, 0, 2, 1)$. The vectors in question are $\mathbf{u} = (2, 1, 5, 4)$ and $\mathbf{w} = (1, 0, 0, 0)$. Which and why?

SOLUTION. Theorem 3.6.1 tells us that if $A = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$, then we need only determine whether or not the systems $A\mathbf{x} = \mathbf{u}$ and $A\mathbf{x} = \mathbf{w}$ are consistent. In the interests of efficiency, we may as well do both at once by forming the augmented matrix for both right hand sides at once as

$$[A \,|\, \mathbf{u} \,|\, \mathbf{v}] = \begin{bmatrix} 1 & 0 & 1 & 2 & 1 \\ 1 & 2 & 0 & 1 & 0 \\ 3 & 2 & 2 & 5 & 0 \\ 3 & 4 & 1 & 4 & 0 \end{bmatrix}$$

The reduced row echelon form of this matrix (whose calculation we leave as an exercise) is

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

Observe that there is a pivot in the fifth column but not in the fourth column. This tells us that the system with augmented matrix $[A \mid \mathbf{u}]$ is consistent, but the system with augmented matrix $[A \mid \mathbf{v}]$ is not consistent. Therefore $\mathbf{u} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, but $\mathbf{v} \notin \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. As a matter of fact, the reduced row echelon form of $[A \mid \mathbf{u}]$ tells us what linear combinations will work, namely

$$\mathbf{u} = (2 - c_3)\mathbf{v}_1 - \frac{1}{2}(1 - c_3)\mathbf{v}_2 + c_3\mathbf{v}_3.$$

where c_3 is an arbitrary scalar. The reason for the non-uniqueness of the coordinates of $\mathbf{u}$ is that the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are not linearly independent. □

The next matter of business is a description of the solution space itself, given that it is not empty. We already have a pretty good conceptual model for the solution of a homogeneous system $A\mathbf{x} = 0$. Remember that this is just the null space, $\mathcal{N}(A)$, of the matrix A. In fact, the definition of $\mathcal{N}(A)$ is the set of vectors $\mathbf{x}$ such that $A\mathbf{x} = 0$. The important point here is that we proved that $\mathcal{N}(A)$ really is a subspace of the appropriate n dimensional standard space $\mathbb{R}^n$ or $\mathbb{C}^n$. As such we can really picture it when n is 2 or 3: $\mathcal{N}(A)$ is either the origin, a line through the origin, a plane through the origin, or in the case $A = 0$, all of $\mathbb{R}^3$. What can we say about a non-homogeneous system? Here is a handy way of understanding these solution sets.

Form of General Solution

THEOREM 3.6.3. *Suppose the system $A\mathbf{x} = \mathbf{b}$ is consistent with a particular solution $\mathbf{x}_0$. Then the general solution $\mathbf{x}$ to this system can be described by the equation*

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{z}$$

where $\mathbf{z}$ runs over all elements of $\mathcal{N}(A)$.

PROOF. On the one hand, suppose we are given a vector of the form $\mathbf{x} = \mathbf{x}_0 + \mathbf{z}$, where $A\mathbf{x}_0 = \mathbf{b}$ and $\mathbf{z} \in \mathcal{N}(A)$. Then

$$A\mathbf{x} = A(\mathbf{x}_0 + \mathbf{z}) = A\mathbf{x}_0 + A\mathbf{z} = \mathbf{b} + 0 = \mathbf{b}.$$

Thus $\mathbf{x}$ is a solution to the system. Conversely, suppose we are given any solution $\mathbf{x}$ to the system and that $\mathbf{x}_0$ is a particular solution to the system. Then

$$A(\mathbf{x} - \mathbf{x}_0) = A\mathbf{x} - A\mathbf{x}_0 = \mathbf{b} - \mathbf{b} = 0$$

It follows that $\mathbf{x} - \mathbf{x}_0 = \mathbf{z} \in \mathcal{N}(A)$ so that $\mathbf{x}$ has the required form $\mathbf{x}_0 + \mathbf{z}$. □

This is really a pretty fact, so let's be clear about what it is telling us. It says that the solution space to a consistent system, as a set, can be described as the set of all translates of elements in the null space of A by some fixed vector. Such a set is sometimes called an *affine set* or a *flat*. When n is 2 or 3 this says that the solution set is either a single point, a line or a plane – *not* necessarily through the origin!

EXAMPLE 3.6.4. Describe geometrically the solution sets to the system

$$\begin{aligned} x+2y &= 3 \\ x+y+z &= 3 \end{aligned}$$

SOLUTION. First solve the system, which has augmented matrix

$$\begin{bmatrix} 1 & 2 & 0 & 3 \\ 1 & 1 & 1 & 3 \end{bmatrix} \xrightarrow[E_{21}(-1)]{} \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & -1 & 1 & 0 \end{bmatrix} \xrightarrow[E_2(-1)]{E_{12}(2)} \begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

The general solution to the system is given in terms of the free variable z, which we will relabel as $z = t$ to obtain

$$\begin{aligned} x &= 3 - 2t \\ y &= t \\ z &= t \end{aligned}$$

We recognize this from calculus as a parametric representation of a line in three dimensional space $\mathbb{R}^3$. Notice that this line does not pass through the origin since $z = 0$ forces $x = 3$. So the solution set is definitely not a subspace of $\mathbb{R}^3$. $\square$

Now we turn to another computational matter. How do we find bases of vector spaces prescribed by a spanning set? How do we find the linear dependencies in a spanning set or implement the Steinitz substitution principle in a practical way? We have all the tools we need now to solve these problems. Let's begin with the question of finding a basis. We are going to solve this problem in two ways. Each has its own merits.

First we examine the row space approach. We require two simple facts.

THEOREM 3.6.5. *Let A be any matrix and E an elementary matrix. Then*

$$\mathcal{R}(A) = \mathcal{R}(EA).$$

PROOF. Suppose the rows of A are the vectors $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n$, so that we have $\mathcal{R}(A) = \operatorname{span}\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n\}$. If $E = E_{ij}$, then the effect of multiplication by E is to switch the ith and jth rows, so the rows of EA are simply the rows of A in a different order. Hence, $\mathcal{R}(A) = \mathcal{R}(EA)$ in this case. If $E = E_i(a)$, with a a nonzero scalar, then the effect of multiplication by E is to replace the ith row by a nonzero multiple of itself. Clearly, this doesn't change the span of the rows either. To simplify notation, consider the case $E = E_{12}(a)$. Then the first row $\mathbf{r}_1$ is replaced by $\mathbf{r}_1 + a\mathbf{r}_2$, so that any combination of the rows of EA is expressible as a linear combination of the rows of A. Conversely, since $\mathbf{r}_1 = \mathbf{r}_1 + a\mathbf{r}_2 - a\mathbf{r}_2$, we see that any combination of $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n$ can be expressed in terms of the rows of EA. This proves the theorem. $\square$

THEOREM 3.6.6. *If the matrix R is in reduced row echelon form, then the nonzero rows of R form a basis of $\mathcal{R}(R)$.*

PROOF. Suppose the rows of R are given as $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n$, so that we have $\mathcal{R}(R) = \operatorname{span}\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k\}$, where the first k rows of R are nonzero and the remaining rows are zero rows. So certainly the nonzero rows span $\mathcal{R}(R)$. In order for these vectors to form a basis, they must also be a linearly independent set. If some linear combination of these vectors is zero, say

$$0 = c_1\mathbf{r}_1 + \cdots + c_k\mathbf{r}_k$$

we examine the ith coordinate of this linear combination, corresponding to the column in which the ith pivot appears. In that column $\mathbf{r}_i$ has a coordinate value of 1 and all other $\mathbf{r}_j$ have a value of zero. Therefore, the linear combination above yields that $c_i = 0$. Since this holds for each $i \leq k$, we obtain that all $c_i = 0$ and the nonzero rows must be linearly independent. It follows that these vectors form a basis of $\mathcal{R}(R)$. □

These theorems are the foundations for the following algorithm for finding a basis for a vector space.

Row Space Algorithm: Given $V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\} \subseteq \mathbb{R}^n$ or $\mathbb{C}^n$.

1. Form the $m \times n$ matrix A whose rows are $\mathbf{v}_1^T, \mathbf{v}_2^T, \dots, \mathbf{v}_m^T$.
2. Find the reduced row echelon form R of A.
3. List the nonzero rows of R. Their transposes form a basis of V.

EXAMPLE 3.6.7. Given that the vector space V is spanned by

$$\begin{aligned} \mathbf{v}_1 &= (1,1,3,3) \\ \mathbf{v}_2 &= (0,2,2,4) \\ \mathbf{v}_3 &= (1,0,2,1) \\ \mathbf{v}_4 &= (2,1,5,4) \end{aligned}$$

Find a basis of V by the row space algorithm.

SOLUTION. Form the matrix

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

Now find the reduced row echelon form of A:

$$\begin{bmatrix} 1 & 1 & 3 & 3 \\ 0 & 2 & 2 & 4 \\ 1 & 0 & 2 & 1 \\ 2 & 1 & 5 & 4 \end{bmatrix} \xrightarrow[\substack{E_{41}(-2) \\ E_2(1/2)}]{E_{31}(-1)} \begin{bmatrix} 1 & 1 & 3 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & -1 & -1 & -2 \\ 0 & -1 & -1 & -2 \end{bmatrix} \xrightarrow[\substack{E_{42}(1) \\ E_{12}(-1)}]{E_{32}(1)} \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

From this we see that the vectors $(1,0,2,1)$ and $(0,1,1,2)$ form a basis for the row space of A.

□

The second algorithm for computing a basis does a little more than find a basis: it gives us a way to tackle the question of what linear combinations sum to zero.

THEOREM 3.6.8. *Let $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ be a matrix with columns $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. Suppose the indices of the non-pivot columns in the reduced row echelon form of A are $i_1, i_2, \dots, i_k$. Then every trivial linear combination*

$$0 = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n$$

of the columns of A is uniquely determined by the values of $c_{i_1}, c_{i_2}, \dots, c_{i_k}$. In particular, if these coefficients are 0, then all the other coefficients must be 0.

PROOF. Express the linear combination in the form

$$0 = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n = A\mathbf{c}$$

where $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ and $A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$. In other words, the column $\mathbf{c}$ of coefficients is in the null space of A. Every solution $\mathbf{c}$ to this system is uniquely specified as follows: assign arbitrary values to the free variables, then use the rows of the reduced row echelon form of A to solve for each bound variable. This is exactly what we wanted to show. □

In view of this theorem, we see that the columns of A corresponding to pivot columns (equivalently, bound variables) in the reduced row echelon form of A must be themselves a linearly independent set. We also see from the proof of this theorem that we can express any column corresponding to a non-pivot column (equivalently, free variable) in terms of columns corresponding to bound variables by setting the free variable corresponding to this column to 1, and all other free variables to 0. Therefore, the columns of A corresponding to pivot columns form a basis of $\mathcal{C}(A)$. This justifies the following algorithm for finding a basis for a vector space.

Column Space Algorithm: Given $V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subseteq \mathbb{R}^m$ or $\mathbb{C}^m$.

1. Form the $m \times n$ matrix A whose columns are $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.
2. Find the reduced row echelon form R of A.
3. List the columns of A that correspond to pivot columns of R. These form a basis of V.

Caution: It is not the columns (nor the rows) of the reduced row echelon form matrix R that yield the basis vectors for V. In fact, if E is an elementary matrix, in general we have $\mathcal{C}(A) \neq \mathcal{C}(EA)$.

EXAMPLE 3.6.9. Given that the vector space V is spanned by

$$\begin{aligned}
\mathbf{v}_1 &= (1, 1, 3, 3) \\
\mathbf{v}_2 &= (0, 2, 2, 4) \\
\mathbf{v}_3 &= (1, 0, 2, 1) \\
\mathbf{v}_4 &= (2, 1, 5, 4)
\end{aligned}$$

Find a basis of V by the column space algorithm.

SOLUTION. Form the matrix A whose columns are these vectors:

$$\begin{bmatrix} 1 & 0 & 1 & 2 \\ 1 & 2 & 0 & 1 \\ 3 & 2 & 2 & 5 \\ 3 & 4 & 1 & 4 \end{bmatrix} \xrightarrow[\substack{E_{21}(-1) \\ E_{31}(-3) \\ E_{41}(-3)}]{} \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 2 & -1 & -1 \\ 0 & 2 & -1 & -1 \\ 0 & 4 & -2 & -2 \end{bmatrix} \xrightarrow[\substack{E_{32}(-1) \\ E_{42}(-2) \\ E_2(1/2)}]{} \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We can see from this calculation that the first and second columns will be pivot columns, while the third and fourth will not be. According to the column space algorithm, $\mathcal{C}(A)$ is a two dimensional space with the first two columns for a basis. □

Just for the record, let's notice here that Theorem 3.6.8 shows us exactly how to express the last two columns in terms of the first two. From the first two rows of the reduced row echelon form of A we see that if $\mathbf{c} = (c_1, c_2, c_3, c_4)$ and $A\mathbf{c} = 0$, then

$$\begin{aligned} c_1 + c_3 + 2c_4 &= 0 \\ c_2 - \frac{1}{2}c_3 - \frac{1}{2}c_4 &= 0 \end{aligned}$$

So for $\mathbf{v}_3$ we choose $c_3 = 1$ and $c_4 = 0$ to obtain that $c_1 = -1$ and $c_2 = 1/2$. Therefore we have

$$-1\mathbf{v}_1 + \frac{1}{2}\mathbf{v}_2 + 1\mathbf{v}_3 + 0\mathbf{v}_4 = 0$$

so that

$$\mathbf{v}_3 = \mathbf{v}_1 - \frac{1}{2}\mathbf{v}_2.$$

A similar calculation with $c_3 = 0$ and $c_4 = 1$ yields that $c_1 = -2$ and $c_2 = 1/2$. Therefore we obtain

$$-2\mathbf{v}_1 + \frac{1}{2}\mathbf{v}_2 + 0\mathbf{v}_3 + 1\mathbf{v}_4 = 0$$

so that

$$\mathbf{v}_4 = 2\mathbf{v}_1 - \frac{1}{2}\mathbf{v}_2$$

Finally, we consider the problem of finding a basis for a null space. Actually, we have already dealt with this problem in an earlier example (Example 3.4.7), but now we will justify what we did there.

THEOREM 3.6.10. *Let A be an $m \times n$ matrix such that $\operatorname{rank} A = r$. Suppose the general solution to the homogeneous equation $A\mathbf{x} = 0$ with $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is written in the form*

$$\mathbf{x} = x_{i_1}\mathbf{v}_1 + x_{i_2}\mathbf{v}_2 + \cdots + x_{i_{n-r}}\mathbf{v}_{n-r}$$

where $x_{i_1}, x_{i_2}, \ldots, x_{i_{n-r}}$ are the free variables. Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-r}$ form a basis of $\mathcal{N}(A)$.

PROOF. The vector $\mathbf{x} = 0$ occurs precisely when all the free variables are set equal to 0, for the bound variables are linear combinations of the free variables. This means that the only linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-r}$ that sum to 0 are those for which all the coefficients $x_{i_1}, x_{i_2}, \ldots, x_{i_{n-r}}$ are 0. Hence these vectors are linearly independent. They span $\mathcal{N}(A)$ since every element $\mathbf{x} \in \mathcal{N}(A)$ is a linear combination of them. Therefore, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-r}$ form a basis of $\mathcal{N}(A)$. $\square$

We see from the statement of this theorem that the nullity of a matrix is simply the dimension of the null space of the matrix. It is also the basis of this algorithm.

Null Space Algorithm: Given an $m \times n$ matrix A.

1. Compute the reduced row echelon form R of A.
2. Use R to find the general solution to the homogeneous system $A\mathbf{x} = 0$.
3. Write the general solution $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ to the homogeneous system in the form
$$\mathbf{x} = x_{i_1}\mathbf{v}_1 + x_{i_2}\mathbf{v}_2 + \cdots + x_{i_{n-r}}\mathbf{v}_{n-r}$$
where $x_{i_1}, x_{i_2}, \ldots, x_{i_{n-r}}$ are the free variables.
4. List the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-r}$. These form a basis of $\mathcal{N}(A)$.

EXAMPLE 3.6.11. Find a basis for the null space of the matrix A in the preceding example by the null space algorithm.

SOLUTION. We already found the reduced row echelon form of A as
$$R = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
The variables x_3 and x_4 are free, while x_1 and x_2 are bound. Hence the general solution of $A\mathbf{x} = 0$ can be written as
$$\begin{aligned} x_1 &= x_3 + 2x_4 \\ x_2 &= -\frac{1}{2}x_3 + -\frac{1}{2}x_4 \\ x_3 &= x_3 \\ x_4 &= x_4 \end{aligned}$$
which becomes, in vector notation,
$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ -1/2 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ -1/2 \\ 0 \\ 1 \end{bmatrix}.$$
Therefore, $\mathbf{v}_1 = (1, -1/2, 1, 0)$ and $\mathbf{v}_1 = (2, -1/2, 0, 1)$ form a basis of $\mathcal{N}(A)$. □

Here is a summary of the key dimensional facts that we have learned in this section:

THEOREM 3.6.12. *Let A be an $m \times n$ matrix such that $\operatorname{rank} A = r$. Then* Rank Theorem

1. $\dim \mathcal{C}(A) = r$
2. $\dim \mathcal{R}(A) = r$
3. $\dim \mathcal{N}(A) = n - r$

3.6 Exercises

1. Find bases for the row space, column space and null space of the following matrices and the dimension of each of these subspaces.

(a) $A = \begin{bmatrix} 0 & 2 \\ -1 & 1 \\ 1 & 1 \end{bmatrix}$ (b) $B = \begin{bmatrix} 2 & 2 & -4 \\ -1 & 0 & 2 \\ 1 & 1 & -2 \end{bmatrix}$ (c) $C = \begin{bmatrix} 1 & 2 & -1 & 2 \\ -1 & 0 & 2 & 2 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ (e) $\begin{bmatrix} 2 & 2 & -4 \\ -4 & -4 & 8 \end{bmatrix}$ (f) $\begin{bmatrix} 1+i & 2 & 2-i \\ -1 & 0 & i \end{bmatrix}$

2. Find all possible linear combinations of the following sets of vectors that sum to $\mathbf{0}$ and the dimension of the subspaces spanned by these vectors.

(a) $(0,1,1)$, $(2,0,1)$, $(2,2,3)$, $(0,2,2)$ in $\mathbb{R}^3$.

(b) x, x^2+x, x^2-x in $\mathcal{P}_2$.

3. Let

$$A = \begin{bmatrix} 4 & 3 & 5 \\ 5 & 4 & 3 \\ 2 & 1 & 9 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 1 & 3 \\ -2 & -1 & -4 \\ 7 & 5 & 17 \end{bmatrix}$$

Compute a basis for (a) $\mathcal{R}(A) + \mathcal{R}(B)$ and for (b) $\mathcal{N}(A) + \mathcal{N}(B)$.

4. In this exercise you should use the fact that B is the reduced row-echelon form of A where

$$A = \begin{bmatrix} 3 & 1 & -2 & 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & -1 & 1 & 2 & 2 \\ 3 & 2 & -1 & 1 & 1 & 8 & 9 \\ 0 & 2 & 2 & -1 & 1 & 6 & 8 \\ 0 & 3 & 3 & 3 & -3 & 0 & 3 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & -2 & -3 \\ 0 & 1 & 1 & 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 1 & 0 & 4 & 5 \\ 0 & 0 & 0 & 0 & 1 & 6 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(a) Find a basis for the row space of A and give the rank of A.

(b) Find a basis for the column space of A.

(c) Find a basis for the set of solutions to $A\mathbf{x} = \mathbf{0}$ and give the nullity of A.

5. Describe geometrically the solution sets to the systems

$$\text{(a)} \quad \begin{aligned} 3x + 6y + 3z &= 9 \end{aligned} \quad \text{(b)} \quad \begin{aligned} x + 2y + z &= 3 \\ x - z &= 1 \end{aligned} \quad \text{(c)} \quad \begin{aligned} 6x + 4y - 4z &= 0 \\ 5x + 3y + 3z &= 6 \\ 6x + 2y - 2z &= 0 \end{aligned}$$

6. Let A be an $m \times n$ matrix of rank r. Suppose that there exists a vector $\mathbf{b} \in R^m$ such that the system $A\mathbf{x} = \mathbf{b}$ is inconsistent. Use the consistency and rank theorems of this section to deduce that the system $A^T\mathbf{y} = \mathbf{0}$ must have nontrivial solutions. *Hint:* What does $b \notin \mathcal{C}(A)$ tell you about r?

7. Find bases for the subspace $V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ of $\mathbb{R}^3$ by the row space algorithm and the column space algorithm, where $\mathbf{v}_1 = (1,1,3,3)$, $\mathbf{v}_2 = (0,2,2,4)$, $\mathbf{v}_3 = (1,0,2,1)$, and $\mathbf{v}_4 = (2,1,5,4)$.

8. Find bases for the subspace $V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ of $\mathbb{R}^3$ by the row space algorithm and the column space algorithm, where $\mathbf{v}_1 = (1, 1, 2, 2), \mathbf{v}_2 = (0, 2, 0, 2)$, $\mathbf{v}_3 = (1, 0, 2, 1)$ and $\mathbf{v}_4 = (2, 1, 4, 3)$.

9. Find two bases for the subspace $V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5\}$ of $\mathcal{P}_2$ where $\mathbf{v}_1 = 1 + x$, $\mathbf{v}_2 = 1 + x - x^2$, $\mathbf{v}_3 = 1 + x + x^2$, $\mathbf{v}_4 = x - x^2$, and $\mathbf{v}_5 = 1 + 2x$. *Hint:* You can tackle this directly or use standard vectors instead, which can be done by the isomorphism of Page 158.

10. Suppose that the linear system $A\mathbf{x} = \mathbf{b}$ is a consistent system of equations, where A is an $m \times n$ matrix and $\mathbf{x} = [x_1, \ldots, x_n]^T$. Prove that if the set of columns of A has redundant vectors in it, then the system has more than one solution.

11. Let $p(x) = c_0 + c_1 x + \cdots + c_m x^m$ be a polynomial and A an $n \times n$ matrix. The value of p at A is defined to be the $n \times n$ matrix

$$p(A) = c_0 I + c_1 A + \cdots + c_m A^m$$

Use the result of Exercise 8 of Section 6 to show that there exists a polynomial $p(x)$ of degree at most n^2 for which $p(A) = 0$. (Aside: this estimate is actually much too pessimistic. There is a theorem, the Cayley-Hamilton theorem, that shows that n works.)

12. Use Theorem 3.6.3 and the Dimension theorem to answer the question posed in Example 3.5 of Section 3.6.

13. Use the rank theorem to prove that any rank 1 matrix can be expressed in the form $\mathbf{u}\mathbf{v}^T$ for suitable standard vectors $\mathbf{u}$ and $\mathbf{v}$.

14. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be bases of U and V, respectively, where U and V are subspaces of the vector space W.

(a) Show that the set of mn vectors $\mathbf{u}_j + \mathbf{v}_k$, $j = 1, 2, \ldots, m$, $k = 1, 2, \ldots, n$ spans the subspace $U + V$.

(b) Show that if $U \cap V = \{\mathbf{0}\}$, then the vectors in (a) form a basis of $U + V$.

(c) Show by example that part (b) fails if $U \cap V \neq \{\mathbf{0}\}$.

3.7. *Change of Basis and Linear Operators

How much information do we need to uniquely identify an operator? For a general operator the answer is: a lot! Specifically, we don't really know everything about it until we know how to find its value at every possible argument. This is an infinite amount of information. Yet we know that in some circumstances we can do better. For example, to know a polynomial function completely, we only need a finite amount of data, namely the coefficients of the polynomial. We have already seen that linear operators are special. Are they described by a finite amount of data? The answer is a resounding "yes" in the situation where the domain and target are finite dimensional.

Let $T : V \to W$ be such an operator. Suppose that $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of V and $C = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ is a basis of W. Now let $\mathbf{v} \in V$ be given. We know that there exists a unique set of scalars (the coordinates of $\mathbf{v}$ with respect to this basis) $c_1, c_2, \ldots, c_n$ such that

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

Thus by linearity of T we see that

$$\begin{aligned} T(\mathbf{v}) &= T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) \\ &= c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_nT(\mathbf{v}_n) \end{aligned}$$

It follows that we know everything about the linear operator T if we know the vectors $T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)$.

Now go a step further. Each vector $T(\mathbf{v}_j)$ can be expressed uniquely as a linear combination of $\mathbf{w}_1, \mathbf{w}_{2,} \ldots, \mathbf{w}_m$, namely

(3.7.1) $$T(\mathbf{v}_j) = a_{1,j}\mathbf{w}_1 + a_{2,j}\mathbf{w}_2 + \cdots + a_{m,j}\mathbf{w}_m.$$

In other words, the scalars $a_{1,j}, a_{2,j}, \ldots a_{m,j}$ are the coordinates of $T(\mathbf{v}_j)$ with respect to the basis $\mathbf{w}_1, \mathbf{w}_{2,} \ldots, \mathbf{w}_m$. Stack these in columns and we now have the $m \times n$ matrix $A = [a_{i,j}]$ which contains everything we need to know in order to compute $T(\mathbf{v})$. In fact, with the above terminology, we have

$$\begin{aligned} T(\mathbf{v}) &= c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_nT(\mathbf{v}_n) \\ &= c_1\left(a_{1,1}\mathbf{w}_1 + a_{2,1}\mathbf{w}_2 + \cdots + a_{m,1}\mathbf{w}_m\right) + \\ &\qquad \cdots + c_n(a_{1,n}\mathbf{w}_1 + a_{2,n}\mathbf{w}_2 + \cdots + a_{m,n}\mathbf{w}_m) \\ &= (a_{1,1}c_1 + a_{1,2}c_2 + \cdots + a_{1,n}c_n)\mathbf{w}_1 + \\ &\qquad \cdots + (a_{m,1}c_1 + a_{m,2}c_2 + \cdots + a_{m,n}c_n)\,\mathbf{w}_m \end{aligned}$$

Look closely and we see that the coefficients of these vectors are themselves coordinates of a matrix product, namely the matrix A times the column vector of coordinates of $\mathbf{v}$ with respect to the chosen basis of V. The result of this matrix multiplication is a column vector whose entries are the coordinates of $T(\mathbf{v})$ relative to the chosen basis of W. So in a certain sense, computing the value of a linear operator amounts to no more than multiplying a (coordinate) vector by the matrix A. Now we make the following definition.

Matrix of Linear Operator

DEFINITION 3.7.1. The *matrix of the linear operator* $T : V \to W$ *relative to the bases* $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of V and $C = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ of W is the matrix $[T]_{B,C} = [a_{i,j}]$ whose entries are specified by Equation 3.7.1. In the case that $B = C$, we simply write $[T]_B$.

Denote the coordinate vector of a vector $\mathbf{v}$ with respect to a basis B by $[\mathbf{v}]_B$. Then the above calculation of $T(\mathbf{v})$ can be stated succinctly in matrix/vector terms as

(3.7.2) $$[T(\mathbf{v})]_C = [T]_{B,C}[\mathbf{v}]_B$$

Even in the case of an operator as simple as the identity map $I(\mathbf{v}) = \mathbf{v}$, the matrix of a linear operator can be useful and interesting.

DEFINITION 3.7.2. Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $C = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ both be bases of V. Then the matrix $[I]_{B,C}$ of the identity map $I : V \to V$ relative to these bases is called the *change of basis* matrix from the basis C to the basis B.

A very fundamental fact about these change of bases matrices is the answer to the following question. Suppose that $T : V \to W$ and $S : U \to V$ are linear operators. Can we relate the matrices of T, S and $T \circ S$? The answer is as follows.

Change of Basis

THEOREM 3.7.3. *Suppose that $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, $C = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$, and $D = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ are bases of V, W and U, respectively, and that $T : V \to W$ and $S : U \to V$ are linear operators. Then*

$$[T \circ S]_{D,C} = [T]_{B,C}[S]_{D,B}$$

An immediate corollary is that if T is invertible, then so is the matrix of T with respect to any basis. In particular, all change of bases matrices are invertible.

We can now also see exactly what happens when we make a change of basis in the domain and target of a linear operator and recalculate the matrix of the operator. Specifically, suppose that $T : V \to W$ and that B, B' are bases of V and C, C' are bases of W. Let P and Q be the change of basis matrices from B' to B and C' to C, respectively. Identify a matrix with its operator action by multiplication and we have a chain of operator maps

$$B' \xrightarrow{P} B \xrightarrow{T} C \xrightarrow{Q^{-1}} C'$$

so that application of the theorem shows that

$$[T]_{B',C'} = Q^{-1}[T]_{B,C}P$$

NOTATION 3.7.4. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator, then the matrix of T with respect to the standard bases of the domain and target of T, which we simply denote as $[T]$ is called the *standard matrix* of T.

We have just obtained a very important insight into the matrix of a linear transformation. Here is the form it takes for the standard spaces.

COROLLARY 3.7.5. *Let that $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear operator, B a basis of $\mathbb{R}^n$ and C a basis of $\mathbb{R}^m$. Let P and Q be the change of basis matrices from the standard bases to B and C, respectively. If A is the matrix of T with respect to the standard bases and M the matrix of T with respect to the bases B and C, then*

$$M = Q^{-1}AP$$

EXAMPLE 3.7.6. Given the linear operator $T : \mathbb{R}^4 \to \mathbb{R}^2$ by the rule

$$T(x_1, x_2, x_3, x_4) = \begin{bmatrix} x_1 + 3x_2 - x_3 \\ 2x_1 + x_2 - x_4 \end{bmatrix}$$

find the standard matrix of T.

SOLUTION. We see that

$$T(\mathbf{e}_1) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, T(\mathbf{e}_2) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, T(\mathbf{e}_3) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, T(\mathbf{e}_4) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Thus, since the standard coordinate vector of a standard vector is simply itself, we have

$$[T] = \begin{bmatrix} 1 & 3 & -1 & 0 \\ 2 & 1 & 0 & -1 \end{bmatrix}$$

□

EXAMPLE 3.7.7. With T as above, find the matrix of T with respect to the domain basis $B = \{(1,0,0,0),(1,1,0,0),(1,0,1,0),(1,0,0,1)\}$ and range basis $C = \{(1,1),(1,-1)\}$.

SOLUTION. Let A be the matrix of the previous example, so it represents the standard matrix of T. Let $B' = \{(1,0,0,0),(1,0,0,0),(0,0,1,0),(0,0,0,1)\}$ and $C' = \{(1,0),(0,1)\}$ be the standard bases for the domain and target of T. Then we have

$$A = [T] = [T]_{B',C'}$$

The change of basis matrix from any basis B to the standard basis is easy to calculate: simply form the matrix that has the vectors of B listed as its columns. In our case, this means that

$$P = [I]_{B,B'} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$Q = [I]_{C,C'} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Now the chain of operators

$$B \overset{P}{\to} B' \overset{A}{\to} C' \overset{Q^{-1}}{\to} C$$

Therefore

$$\begin{aligned} [T]_{B,C} &= Q^{-1}AP \\ &= -\frac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 3 & -1 & 0 \\ 2 & 1 & 0 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{3}{2} & \frac{7}{2} & 1 & 1 \\ -\frac{1}{2} & \frac{1}{2} & -1 & 0 \end{bmatrix} \end{aligned}$$

□

3.7 Exercises

1. Let the operator $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$f((x,y,z)) = \begin{bmatrix} x+2y \\ x-y \\ y+z \end{bmatrix}$$

Show that T is linear and find the standard matrix for T. Determine bases for each of the spaces associated with the operator (domain, range, image and kernel).

2. Let $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $B' = \{\mathbf{u}'_1, \mathbf{u}'_2\}$ be bases of $\mathbb{R}^2$, where $\mathbf{u}_1 = [2,2]^T$, $\mathbf{u}_2 = [4,-1]^T$, $\mathbf{u}'_1 = [1,3]^T$ and $\mathbf{u}'_2 = [-1,-1]^T$.

(a) Find the change of basis (transition) matrix T from B to B'.

(b) Given that $\mathbf{w} = 3\mathbf{u}_1 + 4\mathbf{u}_2$, use (a) to express $\mathbf{w}$ as a linear combination of $\mathbf{u}'_1$ and $\mathbf{u}'_2$

(c) What is the transition matrix from B to the standard basis $B'' = \{\mathbf{e}_1, \mathbf{e}_2\}$?

3. Let the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by the formula $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 - 4x_2 \\ 2x_1 + x_2 \end{bmatrix}$, and $B' = \{\frac{1}{5}[3,4]^T, \frac{1}{5}[-4,3]\}$. Find the matrix of T with respect to the standard basis $B = \{\mathbf{e}_1, \mathbf{e}_2\}$, the transition matrix from B' to B and use this information to find the matrix of T with respect to the basis B.

4. Define the *determinant* of a linear operator $T : V \to V$ to be the determinant of $[T]_B$, where B is any basis of the finite dimensional vector space V. Show that this definition is independent of the basis B.

3.8. *Computational Notes and Projects

Project Topics

Project: Modeling with Directed Graphs II

We develop the background for this project by way of an example. You might also review the material of page 66.

EXAMPLE 3.8.1. Suppose we have a communications network that connects five nodes which we label $1, 2, 3, 4, 5$. Communications between points are not necessarily two-way. We specify the network by listing ordered pairs (i, j),the meaning of which is that

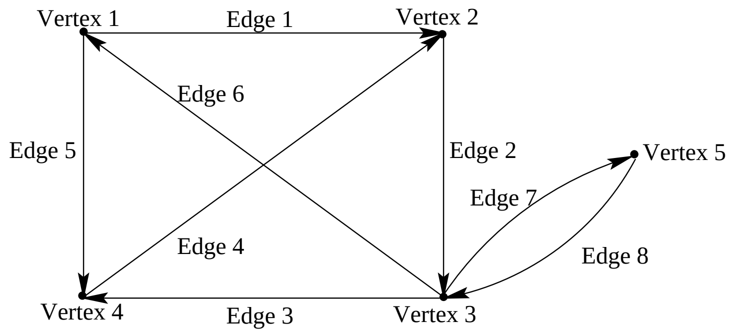


FIGURE 3.8.1. Data from Example 3.8.1

information can be sent from point i to point j. For our problem the connection data is the set

$$E = \{(1,2),(3,1),(1,4),(2,3),(3,4),(3,5),(4,2),(5,3)\}$$

By a *loop* we mean a walk that starts and ends at some node, i.e., a sequence of directed edges connecting a node to itself. For example, the sequence $(3,5),(5,3)$ is a loop in our example. It is important to be able to account for loops in such a network. For one thing, we know that we have two-way communication between any two points in a loop (start at one point and follow the arrows around the loop till you reach the other). Find all the loops of this example and formulate an algorithm that one could program to compute all the loops of the network.

SOLUTION. We recognize this as data that can be modeled by a directed graph (see Example 2.3.4). Thus, "nodes" are just vertices in the graphs and connections are edges. We can draw a picture that contains all the data that we are given by representing each team, or "vertex", as a point and then connecting two points with an arrow, or "directed edge", which points from the winner towards a loser in one of the matches. See Figure 3.8.1 for a picture of this graph.

It isn't so simple to eyeball this graph and count all loops. In fact, if you count going around and around in the same loop as different from the original loop, there are infinitely many. Perhaps we should be a bit more systematic about it. Let's count the smallest loops only, that is, the loops that are not themselves a sum of other loops. It appears that there are only three of these, namely,

$$\begin{aligned} L_1 &: (3,5),(5,3) \\ L_2 &: (2,3),(3,4),(4,2) \\ L_3 &: (1,2),(2,3),(3,1) \end{aligned}$$

There are other loops, e.g., $L_4 : (2,3),(3,5),(5,3),(3,4),(4,2)$. But this is built up out of $L_1$and L_2. In a certain sense, $L_4 = L_1 + L_2$. There seems to be a "calculus of loops." Lurking in the background is another matrix, different from the adjacency matrix that we encountered in Chapter 2, that describes all the data necessary to construct the graph. It is called the *incidence matrix* of the graph and is given as follows: the incidence matrix has its rows index by the vertices of the graph and its columns by the edges. If the edge (i, j) is in the graph, then the column corresponding to this edge has a -1 in its ith row and a $+1$ in its jth row. All other entries are 0. In our example we see that the vertex set is $V = \{1, 2, 3, 4, 5\}$, the edge set is $E = \{(1, 2), (2, 3), (3, 4), (4, 2), (1, 4), (1, 3), (3, 5), (5, 3)\}$ and so the incidence matrix is

$$\begin{aligned} A &= \begin{bmatrix} -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 & \mathbf{v}_5 & \mathbf{v}_6 & \mathbf{v}_7 & \mathbf{v}_8 \end{bmatrix} \end{aligned}$$

We can now describe all loops. Each column of A defines an edge. Thus, linear combinations of these columns with integer coefficients represent a listing of edges, possibly with repeats. Consider such a linear combination with defining vector of coefficients $\mathbf{c} = (c_1, \ldots, c_8)$,

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots c_8\mathbf{v}_8 = A\mathbf{c}$$

When will such a combination represent a loop? For one thing the coefficients should all be nonnegative integers. But this isn't enough. Here's the key idea: we should examine this combination locally, that is, at each vertex. There we expect the total number of "in-arrows" (-1's) to be exactly cancelled by the total number of "out-arrows" ($+1$'s). In other words, each coordinate of $\mathbf{v}$ should be 0 and therefore $\mathbf{c} \in \mathcal{N}(A)$. Now let's find a basis of $\mathcal{N}(A)$ by using the null space algorithm. To make our work a little easier, compute the null space of $-A$ instead of A.

$$\begin{aligned} -A &= \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & -1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \\ &\xrightarrow[\substack{E_{21}(1) \\ E_{32}(1) \\ E_{43}(1)}]{} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \\ &\xrightarrow[\substack{E_{45}(1) \\ E_{34}(-1)}]{} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

From this we see that the free variables are c_4, c_5, c_6, c_8. The general form of an element of the null space takes the form

$$\mathbf{c} = c_4\mathbf{v}_4 + c_5\mathbf{v}_5 + c_6\mathbf{v}_6 + c_8\mathbf{v}_8$$

where the columns are given by setting the corresponding free variable to 1 and the others to 0:

$$\begin{aligned}
\mathbf{v}_4 &= (0, 1, 1, 1, 0, 0, 0, 0) \\
\mathbf{v}_5 &= (-1, -1, -1, 0, 1, 0, 0, 0) \\
\mathbf{v}_6 &= (1, 1, 1, 0, 0, 1, 0, 0) \\
\mathbf{v}_8 &= (0, 0, 0, 0, 0, 0, 1, 1)
\end{aligned}$$

Now we see that $\mathbf{v}_5$ and $\mathbf{v}_6$ don't represent loops in the sense that we originally defined them since we only allow loops to move in the direction of the edge arrow. However, the basis vectors of coefficients that represent loops as we defined them are $\mathbf{v}_4$ and $\mathbf{v}_8$. The loop can be expressed algebraically as $\frac{1}{2}(\mathbf{v}_6 - \mathbf{v}_5)$. Therefore, all possible loops can be represented by the basis vectors for the null space. □

This null space calculation is trying to tell us something. What the null space says is that if we allowed for paths that moved against the direction of the edge arrows when the coefficient of the edge is negative, we would have four independent loops. These "algebraic" loops include our original loops. They are much easier to calculate since we don't have to worry about all the coefficients c_i being of the same sign. They may not be very useful in the context of communication networks, since they don't specify a flow of information; but in the context of electrical circuits they are very important. In fact, the correct definition of a "loop" in electrical circuit theory is an element $\mathcal{N}(A)$ with integer coefficients.

Project Description: This assignment is intended to introduce you to another application of the concept of a graph as used as a tool in mathematical modeling. You are given that the (directed) graph G has vertex set

$$V = \{1, 2, 3, 4, 5, 6\}$$

and edge set

$$E = \{(2, 1), (1, 5), (2, 5), (5, 4), (3, 6), (4, 2), (4, 3), (3, 2), (6, 4), (6, 1)\}$$

Answer the following questions about the graph G. It involves one more idea about graphs. If we thought of the graph as representing an electrical circuit where the presence of an edge indicates some electrical object like a resistor or capacitor we could attach a potential

1. What does the graph look like? You may leave space in your report and draw this by hand or, if you prefer, you may use the computer drawing applications available to you on your system.

2. Next view the graph as representing an electrical circuit with potentials $x_1, ..., x_5$ to be assigned to the vertices. Find $\mathcal{N}(A)$ and $\mathcal{N}(A^T)$ using a computer algebra system available to you. What does the former tell you about the loop structure of the circuit? Distinguish between graphical and "algebraic" loops. Finally, use that fact that $A\mathbf{x} = \mathbf{b}$ implies that for all $\mathbf{y} \in \mathcal{N}(A^T)$, $\mathbf{y}^T\mathbf{b} = 0$ to find conditions that a vector $\mathbf{b}$ must satisfy in order for it to be a vector of potential differences for some potential distribution on the vertices.

Review

Chapter 3 Exercises

1. Use the subspace test to determine which of the following subsets W is a subspace of the vector space V:

(a) V is the space of all 2×2 matrices and W is the subset of matrices of the form $A = \begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$, where a, c are arbitrary scalars and b is an arbitrary nonzero scalar.

(b) V is the space of all polynomials and W is the subset of polynomials that have nonzero constant term.

2. Let $W = \mathbb{R}^{2,2}$ and consider the subsets

$$U = \{ \begin{bmatrix} 0 & 0 \\ a & b \end{bmatrix} \mid a, b \in \mathbb{R}\} \text{ and } V = \{ \begin{bmatrix} 0 & c \\ 0 & d \end{bmatrix} \mid c, d \in \mathbb{R}\}$$

(a) Show that U and V are subspaces of W.

(b) Find a basis for the subspace $U + V$.

(c) Find a basis for the subspace $U \cap V$.

3. Show that $\mathbf{u}_1 = (1, 0, 1)$ and $\mathbf{u}_2 = (1, -1, 1)$ form a linearly independent set. Then fill this set out to a basis of $\mathbb{R}^3$.

4. Show that $1, 1 + x, 1 + x + x^2$ is a basis of $\mathcal{P}_3$ and compute the coordinates of the polynomial $2 - x + 4x^2$ with respect to this basis.

5. Let $T : V \to W$ be a linear operator and suppose that $\dim V = 4$ and $\dim W = 8$. Determine all possible values for $\ker T$ and $\operatorname{range} T$.

6. You are given that $T : V \to \mathbb{R}^3$ is a linear operator, where $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is a basis of V, and moreover $T(\mathbf{v}_1) = (0, 1, 1)$, $T(\mathbf{v}_2) = (1, 1, 0)$ and $T(\mathbf{v}_3) = (-1, 0, 1)$.

(a) Compute $\ker T$. *Hint:* Find conditions on coefficients such that $T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3) = 0$.

(b) Find $\operatorname{range} T$.

(c) Is the vector $(-3, 2, 5)$ in $\operatorname{range} T$? Explain.

7. Let V be a real vector space with basis $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ and define the *coordinate map* as the operator T that assigns to each $v \in V$ the vector of coordinates of v relative to the basis $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. Prove the following:

(a) $T : V \to \mathbb{R}^3$ is a linear operator.

(b) $\ker T = \{0\}$ (c) $\operatorname{range} T = \mathbb{R}^3$.

8. A square matrix $H = [h_{ij}]$ is called *upper Hessenberg* if all entries below the first subdiagonal are zero, that is, $h_{ij} = 0$ when $i > j + 1$. Prove that the set V of all $n \times n$ real Hessenberg matrices is a subspace of $\mathbb{R}^{n,n}$.

9. Answer True/False:

(a) Every spanning set of a vector space contains a basis of the space.

(b) The set consisting of the zero vector is a linearly independent set.

(c) The dimension of the real vector space C^n as a vector space over $\mathbb{R}$ is n.

(d) The vectors $[1, 0]^T$, $[0, 1]^T$ and $[1, 1]^T$ are linearly dependent.

(e) If A is a 6×4 matrix and the system $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions, then A has rank at most 3.

(f) If A is a 4×4 matrix and the system $A\mathbf{x} = \mathbf{b}$ has no solutions, then the columns of A are linearly independent.

(g) In a vector space the set consisting of the zero vector is a linearly independent set.

(h) Every subspace of an infinite dimensional space is infinite dimensional.

(i) A square matrix is invertible if and only if its rows form a linearly independent set.

CHAPTER 4

GEOMETRICAL ASPECTS OF STANDARD SPACES

The standard vector spaces have many important extra features that we have ignored up to this point. These extra features made it possible to do sophisticated calculations in the spaces and enhanced our insight into vector spaces by appealing to geometry. For example, in the geometrical spaces $\mathbb{R}^2$ and $\mathbb{R}^3$ that were studied in calculus, it was possible to compute the length of a vector and angles between vectors. These are visual concepts that feel very comfortable to us. In this chapter we are going to generalize these ideas to the standard spaces and their subspaces. We will abstract these ideas to general vector spaces in Chapter 6.

4.1. Standard Norm and Inner Product

Throughout this chapter vector spaces will be assumed to be subspaces of the standard vector spaces $\mathbb{R}^n$ or $\mathbb{C}^n$.

The Norm Idea

Consider this problem. Can we make sense out of the idea of a sequence of vectors $\mathbf{u}_i$ converging to a limit vector $\mathbf{u}$, i.e.,

$$\lim_{n\to\infty} \mathbf{u}_n = \mathbf{u}$$

in standard spaces? What we need is some idea about the length, or norm, of a vector, so we can say that the length of the difference $\mathbf{u} - \mathbf{u}_n$ should tend to 0 as $n \to \infty$. We have seen such an idea in the geometrical spaces $\mathbb{R}^2$ or $\mathbb{R}^3$. There are different ways to measure length. We shall begin with the most standard method, one which you have already encountered in calculus. It is one of the outcomes of geometry and the Pythagorean theorem. There is no compelling reason to stop at geometrical dimensions of two or three, so here is the general definition.

DEFINITION 4.1.1. Let $\mathbf{u} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. The *(standard) norm* of $\mathbf{u}$ is the nonnegative real number

$$||\,\mathbf{u}\,|| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

EXAMPLE 4.1.2. Compute the norms of the vectors $\mathbf{u} = (1, -1, 3)$ and $\mathbf{v} = [2, -1, 0, 4, 2]^T$.

SOLUTION. From definition

$$||\,\mathbf{u}\,|| = \sqrt{1^2 + (-1)^2 + 3^2} = \sqrt{11} \approx 3.3166$$

and

$$||\,\mathbf{v}\,|| = \sqrt{2^2 + (-1)^2 + 0^2 + 4^2 + 2^2} = \sqrt{25} = 5$$

□

Even though we can't really "see" the five dimensional vector y of this example, it is interesting to note that calculating its length is just as routine as calculating the length of the three dimensional vector x. What about complex vectors? Shouldn't there be an analogous definition of norm for such objects? The answer is "yes," but we have to be a little careful. We can't use the same definition that we did for real vectors. Consider the vector $x = (1, 1 + i)$. The sum of the squares of the coordinates is just

$$1^2 + (1 + i)^2 = 1 + 1 + 2i - 1 = 1 + 2i$$

This isn't good. We don't want "length" to be measured in complex numbers. The fix is very simple. We already have a way of measuring the length of a complex number z, namely the absolute value $|\,z\,|$. So length squared should be $|\,z\,|^2$. That is the inspiration for the following definition which is entirely consistent with our first definition when applied to real vectors:

DEFINITION 4.1.3. Let $\mathbf{u} = (z_1, z_2, \ldots, z_n) \in \mathbb{C}^n$. The *norm* of x is the nonnegative real number

$$||\,\mathbf{u}\,|| = \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2}$$

Notice that $|z|^2 = \overline{z}z$. (Remember that if $z = a + bi$, then $\overline{z} = a - bi$ and $\overline{z}z = a^2 + b^2 = |\,z\,|^2$.) Therefore,

$$||\,\mathbf{u}\,|| = \sqrt{\overline{z}_1 z_1 + \overline{z}_2 z_2 + \cdots + \overline{z}_n z_n}$$

EXAMPLE 4.1.4. Compute the norms of the vectors $\mathbf{u} = (1, 1+i)$ and $\mathbf{v} = [2, -1, i, 3-2i]^T$

SOLUTION. From definition

$$||\,\mathbf{u}\,|| = \sqrt{1^2 + (1 - i)(1 + i)} = \sqrt{1 + 1 + 1} \approx 1.7321$$

and

$$\begin{aligned}||\,\mathbf{v}\,|| &= \sqrt{2^2 + (-1)^2 + (-i)i + (3 + 2i)(3 - 2i)}\\ &= \sqrt{4 + 1 + 1 + 9 + 4} = \sqrt{19} \approx 4.3589\end{aligned}$$

□

Just as we have asked for other key ideas, we now ask the question "What are the essential properties of a norm concept?" The answer:

Basic Norm Laws. Let c be a scalar and $\mathbf{u}, \mathbf{v} \in V$ where the vector space V has the standard norm $||\ \ ||$. Then the following hold.

1. $||\,\mathbf{u}\,|| \geq 0$ with equality if and only if $\mathbf{u} = \mathbf{0}$.
2. $||\,c\mathbf{u}\,|| = |\,c\,|\ ||\,\mathbf{u}\,||$
3. (Triangle Inequality) $||\,\mathbf{u} + \mathbf{v}\,|| \leq ||\,\mathbf{u}\,|| + ||\,\mathbf{v}\,||$

That (1) is true is immediate from the definition of $||\,\mathbf{u}\,||$ as a sum of the lengths squared of the coordinates of $\mathbf{u}$. This sum is zero exactly when each term is zero. Condition (2) is fairly straightforward too. Suppose $\mathbf{u} = (z_1, z_2, \dots, z_n)$, so that

$$\begin{aligned}
||\,c\mathbf{u}\,|| &= \sqrt{(\overline{cz_1})cz_1 + (\overline{cz_2})cz_2 + \cdots + (\overline{cz_n})cz_n} \\
&= \sqrt{(\overline{c}c)(\overline{z}_1 z_1 + \overline{z}_2 z_2 + \cdots + \overline{z}_n z_n)} \\
&= \sqrt{|c|^2}\sqrt{\overline{z}_1 z_1 + \overline{z}_2 z_2 + \cdots + \overline{z}_n z_n} \\
&= |c|\ ||\,\mathbf{u}\,||
\end{aligned}$$

The triangle inequality (which gets its name from the triangle with representatives of the vectors $\mathbf{u}, \mathbf{v}, \mathbf{u}+\mathbf{v}$ as its legs) can be proved easily in two or three dimensional geometrical space by appealing to the fact that the sum of lengths of any two legs of a triangle is greater that the length of the third leg. A justification for higher dimensions is a nontrivial piece of algebra that we will postpone until after the introduction of inner products below.

Unit Vectors

First we consider a few applications of the norm concept. The first of these is the idea of "normalizing" a vector. This means finding a *unit vector*, which means a *vector of length* 1, that has the same direction as the given vector. This process is sometimes called "normalization." How do we do it? The following simple fact helps.

THEOREM 4.1.5. *Let* $\mathbf{u}$ *be a nonzero vector. Then the vector*

$$\mathbf{w} = \frac{1}{||\,\mathbf{u}\,||}\mathbf{u}$$

is a unit vector in the same direction as $\mathbf{u}$.

PROOF. Any two vectors *determine the same direction* if one is a positive multiple of the other (actually, this is a definition of "determining the same direction"). Therefore we see immediately that $\mathbf{w}$ and $\mathbf{u}$ determine the same direction. Now check the length of $\mathbf{w}$ and use basic norm law 2 to obtain that

$$||\,\mathbf{w}\,|| = \left|\frac{1}{||\,\mathbf{u}\,||}\mathbf{u}\right| = \left|\frac{1}{||\,\mathbf{u}\,||}\right| ||\,\mathbf{u}\,|| = \frac{||\,\mathbf{u}\,||}{||\,\mathbf{u}\,||} = 1$$

Hence $\mathbf{w}$ is a unit vector, as desired. □

EXAMPLE 4.1.6. Use the normalization procedure to find unit vectors in the directions of vectors $\mathbf{u} = (2, -1, 0, 4)$ and $\mathbf{v} = (-4, 2, 0, -8)$. Conclude that these vectors determine the same direction.

SOLUTION. Let us find a unit vector in the same direction of each vector. We have parallel

$$||\,\mathbf{u}\,|| = \sqrt{2^2 + (-1)^2 + 0^2 + 4^2} = \sqrt{21}$$

and

$$||\,\mathbf{v}\,|| = \sqrt{-4^2 + (2)^2 + +0^2 + (-8)^2} = \sqrt{84} = 2\sqrt{21}$$

It follows that unit vectors in the directions of $\mathbf{u}$ and $\mathbf{v}$, respectively, are

$$\begin{aligned}
\mathbf{w_1} &= (2, -1, 0, 4)/\sqrt{21} \\
\mathbf{w_2} &= (-4, 2, 0, -8)/(2\sqrt{21}) = (2, -1, 0, 4)/\sqrt{21} = \mathbf{w_1}
\end{aligned}$$

It follows that $\mathbf{u}$ and $\mathbf{v}$ determine the same direction. □

EXAMPLE 4.1.7. Find a unit vector in the direction of the vector $\mathbf{v} = (2+i, 3)$.

SOLUTION. We have

$$\|\mathbf{v}\| = \sqrt{2^2 + (1)^2 + 3^2} = \sqrt{14}$$

It follows that a unit vector in the direction of $\mathbf{v}$ is

$$\mathbf{w} = \frac{1}{\sqrt{14}}(2+i, 3)$$

□

In order to work the next example we must express the idea of vector convergence of a sequence $\mathbf{u}_1, \mathbf{u}_2, \ldots$ to the vector $\mathbf{u}$ in a sensible way. The norm idea makes this straightforward: to say that the $\mathbf{u}_n$'s approach the vector $\mathbf{u}$ should mean that the distance between $\mathbf{u}$ and $\mathbf{u}_n$ goes to 0 as $n \to \infty$. But norm measures distance. Therefore the correct definition is as follows:

DEFINITION 4.1.8. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots$ be a sequence of vectors in the vector space V and $\mathbf{u}$ also a vector in V. We say that the sequence converges to $\mathbf{u}$ and write

Convergence of Vectors

$$\lim_{n\to\infty} \mathbf{u}_n = \mathbf{u}$$

if the sequence of real numbers $\|\mathbf{u}_n - \mathbf{u}\|$ converges to 0, i.e.,

$$\lim_{n\to\infty} \|\mathbf{u}_n - \mathbf{u}\| = 0$$

EXAMPLE 4.1.9. Use the norm concept to justify the statement that sequence of vectors $\mathbf{u}_n$ converges to a limit vector $\mathbf{u}$, i.e.,

$$\lim_{n\to\infty} \mathbf{u}_n = \mathbf{u}$$

where $\mathbf{u}_n = [1 + 1/n^2, 1/(n^2+1), \sin n/n]^T$ and $\mathbf{u} = [1, 0, 0]^T$.

SOLUTION. In our case we have

$$\mathbf{u}_n - \mathbf{u} = \begin{bmatrix} 1 + 1/n^2 \\ 1/(n^2+1) \\ \sin n/n \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/n^2 \\ 1/(n^2+1) \\ \sin n/n \end{bmatrix}$$

so

$$\|\mathbf{u}_n - \mathbf{u}\| = \sqrt{\left(\frac{1}{n}\right)^2 + \left(\frac{1}{(n^2+1)}\right)^2 + \left(\frac{\sin n}{n}\right)^2} \underset{n\to\infty}{\to} \sqrt{0+0+0} = 0$$

which is what we wanted to show. □

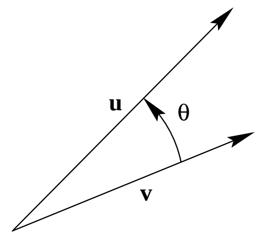


FIGURE 4.1.1. Angle θ between vectors $\mathbf{u}$ and $\mathbf{v}$.

The Inner Product Idea

In addition to norm concept we had another fundamental tool in our arsenal when we tackled two and three dimensional geometrical vectors. This tool was the so-called "dot product" of two vectors. It had many handy applications, but the most powerful of these was the ability to determine the angle between two vectors. In fact, some authors use this idea as the basis for defining dot products as follows: let θ be the angle between representatives of the vectors $\mathbf{u}$ and $\mathbf{v}$. (See Figure 4.1.1.) The dot product of $\mathbf{u}$ and $\mathbf{v}$ is defined to be the quantity $\|\mathbf{u}\| \|\mathbf{v}\| \cos\theta$. It turned out that with some trigonometry and algebra, one could come up with a very convenient form for inner products; for example, in the two dimensional case, if $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$, then

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 \tag{4.1.1}$$

This made the calculation of dot products vastly easier since we didn't have to use any trigonometry to compute it. A particularly nice application was that we could determine $\cos\theta$ quite easily from the dot product, namely

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{||\,\mathbf{u}\,||\;||\,\mathbf{v}\,||} \tag{4.1.2}$$

We have seen that it is useful to try to extend these geometrical ideas to higher dimensions even if we can't literally use trigonometry and the like. So what we do is reverse the sequence of ideas we've discussed and take Equation 4.1.1 as the prototype for our next definition. As with norms, we are going to have to distinguish carefully between the cases of real or complex scalars. First we focus on the more common case of real coefficients.

DEFINITION 4.1.10. Let $\mathbf{u} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{v} = (y_1, y_2, \ldots, y_n)$ be vectors in $\mathbb{R}^n$. The *(standard) inner product*, also called the *dot product* of $\mathbf{u}$ and $\mathbf{v}$, is the real number

$$\begin{aligned} \mathbf{u} \cdot \mathbf{v} &= \mathbf{u}^T \mathbf{v} \\ &= x_1 y_1 + x_2 y_2 + \ldots + x_n y_n \end{aligned}$$

We can see from the first form of this definition where the term "inner product" came from. Recall from Section 2.4 of Chapter 2 that the matrix product $\mathbf{u}^T\mathbf{v}$ is called the inner product of these two vectors.

EXAMPLE 4.1.11. Compute the dot product of the vectors $\mathbf{u} = (1, -1, 3, 2)$ and $\mathbf{v} = (2, -1, 0, 4)$ in $\mathbb{R}^4$.

SOLUTION. From definition

$$\begin{aligned}\mathbf{u}\cdot\mathbf{v} &= 1\cdot 2+(-1)\cdot(-1)+3\cdot 0+2\cdot 4\\ &= 11.\end{aligned}$$

□

There is a wonderful connection between the standard inner product and the standard norm for vectors which is immediately evident from the definitions. Here it is:

(4.1.3)
$$||\,\mathbf{u}\,|| = \sqrt{\mathbf{u}\cdot\mathbf{u}}$$

Thus computing norms amounts to an inner product calculation followed by a square root. Actually, we can even avoid the square root and put the equation in the form

$$||\,\mathbf{u}\,||^2 = \mathbf{u}\cdot\mathbf{u}$$

We say that the standard norm is *induced* by the standard inner product. We would like this property to carry over to complex vectors. Now we have to be a bit careful. In general, the quantity $\mathbf{u}^T\mathbf{u}$ may not even be a real number, or may be negative. This means that $\sqrt{\mathbf{u}^T\mathbf{u}}$ could be complex, which doesn't seem like a good idea for measuring "length." So how can we avoid this problem? Recall that when we introduced transposes, we also introduced Hermitian transposes and remarked that for complex vectors, this is a more natural tool than the transpose. Now we can back up that remark! Recall the definition for complex norm: for $\mathbf{u}=(z_1,z_2,\ldots,z_n)\in\mathbb{C}^n$, the *norm* of x is the nonnegative real number

$$\begin{aligned}||\,\mathbf{u}\,|| &= \sqrt{\overline{z}_1 z_1+\overline{z}_2 z_2+\cdots+\overline{z}_n z_n}\\ &= \sqrt{\mathbf{u}^H\mathbf{u}}.\end{aligned}$$

Therefore, in our definition of complex "dot products" we had better replace transposes by Hermitian transposes. This inspires the definition

DEFINITION 4.1.12. Let $\mathbf{u}=(w_1,w_2,\ldots,w_n)$ and $\mathbf{v}=(z_1,z_2,\ldots,z_n)$ be vectors in $\mathbb{C}^n$. The *(standard) inner product*, which is also called the *dot product* of $\mathbf{u}$ and $\mathbf{v}$, is the complex number

$$\begin{aligned}\mathbf{u}\cdot\mathbf{v} &= \overline{w}_1 z_1+\overline{w}_2 z_2+\ldots+\overline{w}_n z_n\\ &= \mathbf{u}^H\mathbf{v}\end{aligned}$$

With this definition we still have the close connection given above in (4.1.3) between norm and standard inner product of complex vectors.

EXAMPLE 4.1.13. Compute the dot product of the vectors $\mathbf{u}=(1+2i,i,1)$ and $\mathbf{v}=(i,-1-i,0)$ in $\mathbb{C}^3$.

SOLUTION. Just apply the definition:

$$\begin{aligned}\mathbf{u}\cdot\mathbf{v} &= \overline{(1+2i)}i+\overline{i}(-1-i)+1\cdot 0\\ &= (1-2i)i-i(-1-i)\\ &= 1+2i\end{aligned}$$

□

What are the really essential defining properties of these standard inner products? It turns out that we can answer the question for both real and complex inner products at once. However, we should bear in mind that most of the time we will be dealing with real dot products, and in this case all the dot products in questions are real numbers, so that any reference to a complex conjugate can be omitted.

Basic Inner Product Laws. Let c be a scalar and $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ where V is a vector space with the standard inner product. Then the following hold.

1. $\mathbf{u} \cdot \mathbf{u} \geq 0$ with equality if and only if $\mathbf{u} = \mathbf{0}$
2. $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$
3. $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$
4. $\mathbf{u} \cdot (c\mathbf{v}) = c(\mathbf{u} \cdot \mathbf{v})$

That (1) is true is immediate from the fact that $\mathbf{u} \cdot \mathbf{u} = \mathbf{u}^H \mathbf{u}$ is a sum of the length squared of the coordinates of $\mathbf{u}$. This sum is zero exactly when each term is zero. Condition (2) follows from this line of calculation:

$$\overline{\mathbf{v} \cdot \mathbf{u}} = \overline{\mathbf{v}^H \mathbf{u}} = \left(\overline{\mathbf{v}^H \mathbf{u}}\right)^T = (\mathbf{v}^H \mathbf{u})^H = \mathbf{u}^H \mathbf{v} = \mathbf{u} \cdot \mathbf{v}$$

One point that stands out in this calculation is the following

Caution: A key difference between real and complex inner products is in the commutative law $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$, which holds for real vectors but *not* for complex vectors, where $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$.

Conditions (3) and (4) are similarly verified and left to the exercises. We can also use (4) to prove this fact for real vectors:

$$(c\mathbf{u}) \cdot \mathbf{v} = \mathbf{v} \cdot (c\mathbf{u}) = c(\mathbf{v} \cdot \mathbf{u}) = c(\mathbf{u} \cdot \mathbf{v})$$

If we are dealing with complex dot products, matters are a bit trickier. One can show then that

$$(c\mathbf{u}) \cdot \mathbf{v} = \overline{c}(\mathbf{u} \cdot \mathbf{v})$$

so we don't quite have the symmetry that we have for real dots.

4.1 Exercises

1. For the following pairs of vectors, calculate $||\mathbf{u}||$, $\mathbf{u} \cdot \mathbf{v}$, and a unit vector in the direction of $\mathbf{u}$:

(a) $\mathbf{u} = [1, -1]^T$ and $\mathbf{v} = [-2, 3]^T$ (b)$\mathbf{u} = (2, 0, 1)$ and $\mathbf{v} = (-3, 4, 1)$

(c)$\mathbf{u} = [1, 2, 2 - i, 0]^T$ and $\mathbf{v} = [-2, 1, 1, 1]^T$ (d) $\mathbf{u} = (1 + 2i, 2 + i)$ and $\mathbf{v} = (4 + 3i, 1)$.

2. Verify that $\mathbf{u}_n$ converges to a limit vector $\mathbf{u}$, where $\mathbf{u}_n = [2/n, (1+n^2)/(2n^2+n+1)]^T$ by using the norm definition of vector limit.

3. Compute an angle θ between the following pairs of real vectors.

(a) $(3, -5)$ and $(2, 4)$ (b) $(3, 4)$ and $(4, -3)$ (c) $(1, 1, 2)$ and $(2, -1, 3)$

4. Let $c = 3$, $\mathbf{u} = (4, -1, 2, 3)$ and $\mathbf{v} = (-2, 2, -2, 2)$.

(a) Verify that the three basic norm laws hold for these vectors and scalars.

(b) Verify the four basic inner product laws for these vectors and scalars.

5. Verify basic norm law 1: $||\,\mathbf{u}\,|| \geq 0$ with equality if and only if $\mathbf{u} = \mathbf{0}$.

6. Prove that if $\mathbf{v}$ is a vector and c is a positive real, then normalizing $\mathbf{v}$ and normalizing $c\mathbf{v}$ yield the same unit vector. How are the normalized vectors related if c is negative?

7. Show that for real vectors $\mathbf{u}$, $\mathbf{v}$ and real number c one has

$$(c\mathbf{u}) \cdot \mathbf{v} = \mathbf{v} \cdot (c\mathbf{u}) = c(\mathbf{v} \cdot \mathbf{u}) = c(\mathbf{u} \cdot \mathbf{v})$$

8. Show that if $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ (or $\mathbb{C}^n$) and c is a scalar, then

(a) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$

(b) $\mathbf{u} \cdot (c\mathbf{v}) = c(\mathbf{u} \cdot \mathbf{v})$

9. Show from definition that if $\lim_{n\to\infty} \mathbf{u}_n = \mathbf{u}$, where $\mathbf{u}_n = (x_n, y_n) \in \mathbb{R}^2$ and $\mathbf{u} = (x, y)$, then $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} y_n = y$.

10. Show that for any two vectors $\mathbf{u}, \mathbf{v}$ in the same space, $|\|\mathbf{u}\| - \|\mathbf{v}\|| \leq \|\mathbf{u} - \mathbf{w}\|$. *Hint:* Apply the triangle inequality to $\mathbf{u} + (\mathbf{v} - \mathbf{u})$ and $\mathbf{v} + (\mathbf{u} - \mathbf{v})$.

4.2. Applications of Norms and Inner Products

Projections and Angles

Now that we have dot products under our belts we can tackle geometrical issues like angles between vectors in higher dimensions. For the matter of angles, we will stick to real vector spaces, though we could do it for complex vector spaces with a little extra work. What we would like to do is take Equation 4.1.2 as the *definition* of the angle between two vectors. There's one slight problem: how do we know that it will give a quantity that could be a cosine? After all, cosines only take on values between -1 and 1. We could use some help and the Cauchy-Bunyakovsky-Schwarz inequality (CBS for short) is just what we need:

THEOREM 4.2.1. *For vectors* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, CBS Inequality

$$|\,\mathbf{u} \cdot \mathbf{v}\,| \leq ||\,\mathbf{u}\,||\,||\,\mathbf{v}\,||$$

PROOF. Let c be an arbitrary real number and compute the nonnegative quantity

$$\begin{aligned} f(c) &= ||\,\mathbf{u} + c\mathbf{v}\,||^2 \\ &= (\mathbf{u} + c\mathbf{v}) \cdot (\mathbf{u} + c\mathbf{v}) \\ &= \mathbf{u} \cdot \mathbf{u} + \mathbf{u} \cdot (c\mathbf{v}) + (c\mathbf{v}) \cdot \mathbf{u} + (c\mathbf{v}) \cdot (c\mathbf{v}) \\ & ||\,\mathbf{u}\,||^2 + 2c(\mathbf{u} \cdot \mathbf{v}) + c^2||\,\mathbf{v}\,||^2. \end{aligned}$$

The function $f(c)$ is therefore a quadratic in the variable c with nonnegative values. The low point of this quadratic occurs where $f'(c) = 0$, that is, where

$$0 = 2(\mathbf{u} \cdot \mathbf{v}) + 2c||\,\mathbf{v}\,||^2$$

i.e., where

$$c = \frac{-(\mathbf{u} \cdot \mathbf{v})}{||\,\mathbf{v}\,||^2}$$

Evaluate f at this point to get that

$$0 \le ||\,\mathbf{u}\,||^2 - 2\frac{(\mathbf{u} \cdot \mathbf{v})^2}{||\,\mathbf{v}\,||^2} + \frac{(\mathbf{u} \cdot \mathbf{v})^2}{||\,\mathbf{v}\,||^4}||\,\mathbf{v}\,||^2 = ||\,\mathbf{u}\,||^2 - \frac{(\mathbf{u} \cdot \mathbf{v})^2}{||\,\mathbf{v}\,||^2}$$

Now add $(\mathbf{u} \cdot \mathbf{v})^2/||\,\mathbf{v}\,||^2$ to both sides and multiply by $||\,\mathbf{v}\,||^2$ to obtain that

$$(\mathbf{u} \cdot \mathbf{v})^2 \le ||\,\mathbf{u}\,||^2\,||\,\mathbf{v}\,||^2$$

Take square roots and we have the desired inequality. □

This inequality has a number of useful applications. For instance, because of it we can articulate the following definition. There is a certain ambiguity in discussing angle between vectors, since more than one angle works. Actually it's the cosine of these angles that is really unique.

Angle Between Vectors

DEFINITION 4.2.2. For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ we define the *angle* between $\mathbf{u}$ and $\mathbf{v}$ to be any angle θ satisfying

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{||\,\mathbf{u}\,||\;||\,\mathbf{v}\,||}$$

Thanks to the CBS Inequality, we know that $|\,\mathbf{u} \cdot \mathbf{v}\,|/(||\,\mathbf{u}\,||\;||\,\mathbf{v}\,||) \le 1$ so that this formula for $\cos\theta$ makes sense.

EXAMPLE 4.2.3. Find the angle between the vectors $\mathbf{u} = (1, 1, 0, 1)$ and $\mathbf{v} = (1, 1, 1, 1)$ in $\mathbb{R}^4$.

SOLUTION. We have that

$$\cos\theta = \frac{(1,1,0,1) \cdot (1,1,1,1)}{||\,(1,1,0,1)\,||\;||\,(1,1,1,1)\,||} = \frac{3}{2\sqrt{3}} = \frac{\sqrt{3}}{2}.$$

Hence we can take $\theta = \pi/6$. □

EXAMPLE 4.2.4. Use the laws of inner products and the CBS Inequality to verify the triangle inequality for vectors $\mathbf{u}$ and $\mathbf{v}$. What happens to this inequality if we also know that $\mathbf{u} \cdot \mathbf{v} = 0$?

SOLUTION. Here the trick is to avoid square roots. So square both sides of Equation 4.1.3 to obtain that

$$\begin{aligned}||\mathbf{u}+\mathbf{v}||^2 &= (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) \\ &= \mathbf{u}\cdot\mathbf{u}+\mathbf{u}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v} \\ &= ||\mathbf{u}||^2+2(\mathbf{u}\cdot\mathbf{v})+||\mathbf{v}||^2 \\ &\leq ||\mathbf{u}||^2+2\,|\mathbf{u}\cdot\mathbf{v}|+||\mathbf{v}||^2 \\ &\leq ||\mathbf{u}||^2+2\,||\mathbf{u}||\;||\mathbf{v}||+||\mathbf{v}||^2 \\ &= (||\mathbf{u}||+||\mathbf{v}||)^2\end{aligned}$$

where the last inequality follows from the CBS Inequality. If $\mathbf{u}\cdot\mathbf{v}=0$, then the single inequality can be replaced by an equality. □

We have just seen a very important case of angles between vectors that warrants its own name. Recall from geometry that two vectors are *perpendicular* or *orthogonal* if the angle between them is $\pi/2$. Since $\cos\pi/2=0$, we see that this amounts to the equation $\mathbf{u}\cdot\mathbf{v}=0$. Now we can extend the perpendicularity idea to arbitrary vectors, including complex vectors.

DEFINITION 4.2.5. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in the same vector space are *orthogonal* if $\mathbf{u}\cdot\mathbf{v}=0$. In this case we write $\mathbf{u}\perp\mathbf{v}$.

Orthogonal Vectors

In the case that one of the vectors is the zero vector, we have the little oddity that the zero vector is orthogonal to every other vector, since the dot product is always 0 in this case. Some authors require that $\mathbf{u}$ and $\mathbf{v}$ be nonzero as part of the definition. It's a minor point and we won't worry about it. When $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, i.e., $\mathbf{u}\cdot\mathbf{v}=0$, we see from the third equality in the derivation of CBS above that

$$||\mathbf{u}+\mathbf{v}||^2=||\mathbf{u}||^2+||\mathbf{v}||^2$$

which is really the Pythagorean theorem for vectors in $\mathbb{R}^n$.

Pythagorean Theorem

EXAMPLE 4.2.6. Determine if the following pairs of vectors are orthogonal.

(a) $\mathbf{u}=(2,-1,3,1)$ and $\mathbf{v}=(1,2,1,-2)$.

(b) $\mathbf{u}=(1+i,2)$ and $\mathbf{v}=(-2i,1+i)$.

SOLUTION. For (a) we calculate

$$\mathbf{u}\cdot\mathbf{v}=2\cdot 1+(-1)2+3\cdot 1+1(-2)=1$$

so that $\mathbf{u}$ is not orthogonal to $\mathbf{v}$. For (b) we calculate

$$\begin{aligned}\mathbf{u}\cdot\mathbf{v} &= (1-i)(-2i)+2(1+i) \\ &= -2i-2+2+2i=0.\end{aligned}$$

so that $\mathbf{u}$ is orthogonal to $\mathbf{v}$ in this case. □

The next example illustrates a really handy little trick that is well worth remembering.

EXAMPLE 4.2.7. Given a vector (a,b) in $\mathbb{R}^n$ or $\mathbb{C}^n$, find a vector orthogonal to (a,b).

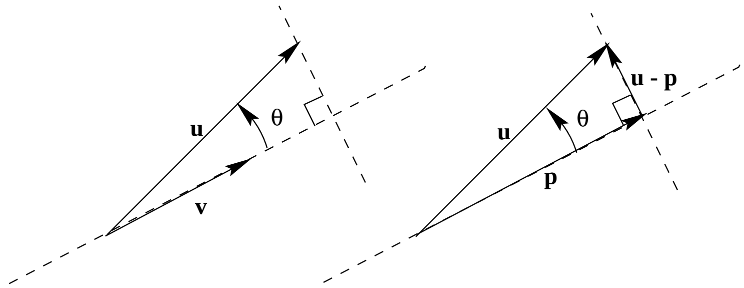


FIGURE 4.2.1. Angle between vectors $\mathbf{u}$ and $\mathbf{v}$, and projection $\mathbf{p}$ of $\mathbf{u}$ along $\mathbf{v}$.

Solution. Simply interchange coordinates, conjugate them (this does nothing if entries are real) and insert a minus sign in front of one of the coordinates, say the first. We obtain $(-\overline{b}, \overline{a})$. Now check that

$$(a, b) \cdot (-\overline{b}, \overline{a}). = -\overline{a}(-\overline{b}) + \overline{b}\overline{a} = 0$$

□

By *parallel vectors* we mean two vectors that are nonzero scalar multiples of each other. Notice that parallel vectors may determine the same or opposite directions. Our next application of the dot product relates back to a fact that we learned in geometry: given two nonzero vectors in the plane, it is always possible to resolve one of them into a sum of a vector parallel to the other and a vector orthogonal to the other (see Figure 4.2.1). The parallel component was called the projection of one vector along the other. As a matter of fact, we can develop this same idea in arbitrary standard vector spaces. That is the content of the following useful fact. Remember, by the way, that "parallel" vectors simply means that the vectors in question are scalar multiples of each other (any scalar).

Projection Formula for Vectors

THEOREM 4.2.8. *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be vectors in a vector space with* $\mathbf{v} \neq 0$. *Let*

$$\mathbf{p} = \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{v} \cdot \mathbf{v}}\mathbf{v} \quad \textit{and} \quad \mathbf{q} = \mathbf{u} - \mathbf{p}$$

Then $\mathbf{p}$ *is parallel to* $\mathbf{v}$, $\mathbf{q}$ *is orthogonal to* $\mathbf{v}$ *and* $\mathbf{u} = \mathbf{p} + \mathbf{q}$.

PROOF. Let $\mathbf{p} = c\mathbf{v}$, an arbitrary multiple of $\mathbf{v}$. Then $\mathbf{p}$ is automatically parallel to $\mathbf{v}$. Impose the constraint that $\mathbf{q} = \mathbf{u} - \mathbf{p}$ be orthogonal to $\mathbf{v}$. This means, by definition, that

$$0 = \mathbf{v} \cdot \mathbf{q} = \mathbf{v} \cdot (\mathbf{u} - \mathbf{p}) = \mathbf{v} \cdot \mathbf{u} - \mathbf{v} \cdot (c\mathbf{v}).$$

Add $\mathbf{v} \cdot (c\mathbf{v})$ to both sides and pull the scalar c outside the dot product to obtain that

$$c(\mathbf{v} \cdot \mathbf{v}) = \mathbf{v} \cdot \mathbf{u}$$

and therefore

$$c = \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{v} \cdot \mathbf{v}}$$

So for this choice of c, $\mathbf{q}$ is orthogonal to $\mathbf{p}$. Clearly, $\mathbf{u} = \mathbf{p} + \mathbf{u} - \mathbf{p}$, so the proof is complete. □

It is customary to call the vector $\mathbf{p}$ of this theorem the *projection of* $\mathbf{u}$ *along* $\mathbf{v}$. We write

Components and Projections

$$\operatorname{proj}_{\mathbf{v}} \mathbf{u} = \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{v} \cdot \mathbf{v}} \mathbf{v}$$

The projection of one vector along another is itself a vector quantity. A *scalar* quantity that is frequently associated with these calculations is the so-called *component* of $\mathbf{u}$ along $\mathbf{v}$. It is defined as

$$\operatorname{comp}_{\mathbf{v}} \mathbf{u} = \frac{\mathbf{v} \cdot \mathbf{u}}{\|\mathbf{v}\|}$$

The connection between these two quantities is that

$$\operatorname{proj}_{\mathbf{v}} \mathbf{u} = \operatorname{comp}_{\mathbf{v}} \mathbf{u} \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

Notice that $\mathbf{v}/\|\mathbf{v}\|$ is a unit vector in the same direction as $\mathbf{v}$. Therefore, $\operatorname{comp}_{\mathbf{v}} \mathbf{u}$ is the signed magnitude of the projection of $\mathbf{u}$ along $\mathbf{v}$.

EXAMPLE 4.2.9. Calculate the projection and component of $\mathbf{u} = (1, -1, 1, 1)$ along the $\mathbf{v} = (0, 1, -2, -1)$ and verify that $\mathbf{u} - \mathbf{p} \perp \mathbf{v}$.

SOLUTION. We have that

$$\begin{aligned}
\mathbf{v} \cdot \mathbf{u} &= 0 \cdot 1 + 1(-1) + (-2)1 + (-1)1 = -4 \\
\mathbf{v} \cdot \mathbf{v} &= 0^2 + 1^2 + (-2)^2 + (-1)^2 = 6
\end{aligned}$$

so that

$$\mathbf{p} = \operatorname{proj}_{\mathbf{v}} \mathbf{u} = \frac{-4}{6}(0, 1, -2, -1) = \frac{1}{3}(0, -2, 4, 2)$$

It follows that

$$\mathbf{u} - \mathbf{p} = \frac{1}{3}(3, -1, -1, 1)$$

and

$$(\mathbf{u} - \mathbf{p}) \cdot \mathbf{v} = \frac{1}{3}(3 \cdot 0 + 1(-1) + (-1)(-2) + 1(-1)) = 0.$$

Also, the component of $\mathbf{u}$ along $\mathbf{v}$ is

$$\operatorname{comp}_{\mathbf{v}} \mathbf{u} = \frac{\mathbf{v} \cdot \mathbf{u}}{\|\mathbf{v}\|} = \frac{-4}{\sqrt{6}}.$$

□

Least Squares

EXAMPLE 4.2.10. You are using a pound scale to measure weights for produce sales when you notice that your scale is about to break. The vendor at the next stall is leaving and loans you another scale as she departs. Soon afterwards your scale breaks. You then realize that the new scale is in units you don't recognize. You happen to have a some known weights that are approximately 2, 5 and 7 pounds respectively. When you weigh these items on the new scale you get the numbers 0.9, 2.4 and 3.2. You get your calculator out and hypothesize that the unit of weight should be some constant multiple of pounds. Model this information as a system of equations. Is it clear from this system what the units of the scale are?

SOLUTION. Express the relationship between the weight p in pounds and the weight w in unknown units as $w \cdot c = p$, where c is an unknown. Your data show that we have

$$\begin{aligned} 0.7c &= 2 \\ 2.4c &= 5 \\ 3.4c &= 7 \end{aligned}$$

As a system of three equations in one unknown you see immediately that this overdetermined system (too many equations) is inconsistent. After all, the pound weights were only approximate and in addition there is always some error in measurement. Consequently, it is not at all clear what the units of the scale are, and we will have to investigate this problem further. You could just average the three inconsistent values of c, thereby obtaining

$$c = (2/0.7 + 5/2.4 + 7/3.4)/3 = 2.3331$$

It isn't at all clear that this should be a good strategy. □

There really is a better way and it will lead to a slightly different estimate of the number c. This method, called the *method of least squares*, was invented by C. F. Gauss to handle uncertainties in orbital calculations in astronomy.

Here is the basic problem: suppose we have data that leads to a system of equations for unknowns that we want to solve for, but the data has errors in it and consequently leads to an inconsistent linear system

$$A\mathbf{x} = \mathbf{b}$$

How do we find the "best" approximate solution? One could answer this in many ways. One of the most commonly accepted ideas is one that goes back to C. F. Gauss: the quantity so-called *residual* $\mathbf{r} = \mathbf{b} - A\mathbf{x}$ should be 0 so its departure from 0 is a measure of our error. Thus we should try to find a value of the unknown $\mathbf{x}$ that minimizes the norm of the residual squared, i.e., a "solution" $\mathbf{x}$ so that

$$||\mathbf{b} - A\mathbf{x}||^2$$

is minimized. Such a solution is called a "least squares" solution to the system. This is used extensively by statisticians, in situations where one has many estimates for unknown parameters which, taken together, are not perfectly consistent. Let's try to get a fix on this problem. Even the 1 variable case is instructive, so let's use the preceding example.

In this case the coefficient matrix A is the column vector $\mathbf{a} = [0.7, 2.4, 3.4]^T$ and the right hand side vector is $\mathbf{b} = [2, 5, 7]^T$. What we are really trying to find is a value of the scalar $x = c$ such that $\mathbf{b} - A\mathbf{x} = \mathbf{b} - x\mathbf{a}$ is a minimum. Here is a geometrical interpretation: we want to find the multiple of the vector $\mathbf{a}$ that is closest to $\mathbf{b}$. Geometry suggests that this minimum occurs when $\mathbf{b} - x\mathbf{a}$ is orthogonal to $\mathbf{a}$, in other words, when $x\mathbf{a}$ is the projection of $\mathbf{b}$ along $\mathbf{a}$. Inspection of the projection formula shows us that we must have

$$x = \frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}} = \frac{0.7 \cdot 2 + 2.4 \cdot 5 + 3.4 \cdot 7}{0.7 \cdot 0.7 + 2.4 \cdot 2.4 + 3.4 \cdot 3.4} = 2.088\,7.$$

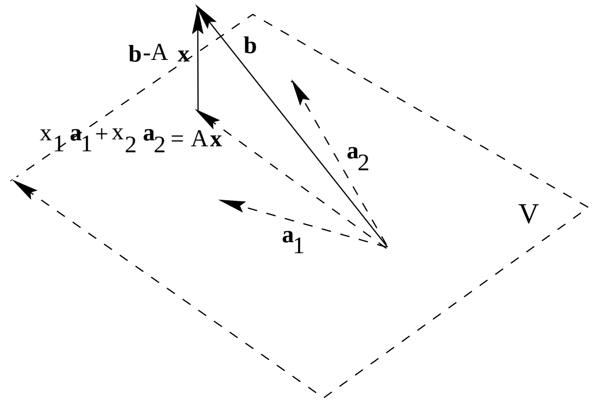


FIGURE 4.2.2. The vector in subspace $\mathcal{C}(A)$ nearest to $\mathbf{b}$.

Notice that this doesn't solve any of the original equations exactly, but it is, in a certain sense, the best approximate solution to all three equations taken together. Also, this solution is *not* the same as the average of the solutions to the three equations, which we computed to be 2.3331.

Now how do we tackle the more general system $A\mathbf{x} = \mathbf{b}$? Since $A\mathbf{x}$ is just a linear combination of the columns, what we should find is the vector of this form which is closest to the vector $\mathbf{b}$. See Figure 6 for a picture of the situation with $n = 2$. Our experience with the 1-dimensional case suggests that we should require that the residual be orthogonal to each column of A, that is, $\mathbf{a}_i \cdot (\mathbf{b} - A\mathbf{x}) = \mathbf{a}_i^T(\mathbf{b} - A\mathbf{x}) = 0$, for all columns $\mathbf{a}_i$ of A. Each column gives rise to one equation. We can write all these equations at once in the form of the so-called *normal equations*: Normal Equations

$$A^T A\mathbf{x} = A^T\mathbf{b}$$

In fact, this is the same set of equations we get if we were to apply calculus to the scalar function of variables $x_1, x_2, ..., x_n$ given as $f(x) = ||\mathbf{b} - A\mathbf{x}||^2$ and search for a local minimum by finding all partials and setting them equal to 0. Any solution to this system will minimize the norm of the difference $\mathbf{b} - A\mathbf{x}$ as $\mathbf{x}$ ranges over all elements of $\mathbb{R}^n$. The coefficient matrix $B = A^T A$ of the normal system has some pleasant properties. For one, it is a symmetric matrix. For another, it is a *positive semidefinite matrix*, by which we mean that B is a square (say $n \times n$) matrix such that $\mathbf{x}^T B\mathbf{x} \geq 0$ for all vectors $\mathbf{x} \in \mathbb{R}^n$. In fact, in some cases B is even better behaved because it is a *positive definite matrix* , by which we mean that B is a square (say $n \times n$) matrix such that $\mathbf{x}^T B\mathbf{x} \geq 0$ for all vectors $\mathbf{x} \in \mathbb{R}^n$. Positive Definite Matrix

Does there exist a solution to the normal equations? The answer is "yes." In general, any solution to the normal equations minimizes the residual norm and is called a *least squares solution* to the problem $A\mathbf{x} = \mathbf{b}$. Since we now have two versions of "solution" for the system $A\mathbf{x} = \mathbf{b}$, we should distinguish between them in situations which may refer to either. If the vector $\mathbf{x}$ actually satisfies the equation $A\mathbf{x} = \mathbf{b}$, we call $\mathbf{x}$ a *genuine solution* to the system to contrast it with a least squares solution. Certainly, every genuine solution is a least squares solution, but the coverse will not be true if the original system is inconsistent. We leave the verifications as exercises.

The normal equations are guaranteed to be consistent – a nontrivial fact – and will have infinitely many solutions if $A^T A$ is a singular matrix. However, we will focus on the most common case, namely that in which A is a rank n matrix. Recall that in this case we say that A has *full column rank*. We can show that the $n \times n$ matrix $A^T A$ is also rank n. This means that it is an invertible matrix and therefore the solution to the normal equations is *unique*. Here is the necessary fact.

THEOREM 4.2.11. *Suppose that the $m \times n$ matrix A has full column rank n. Then the $n \times n$ matrix $A^T A$ also has rank n and is invertible.*

PROOF. Assume A has rank n. Now suppose that for some vector $\mathbf{x}$ we have

$$0 = A^T A\mathbf{x}$$

Multiply on the left by $\mathbf{x}^T$ to obtain that

$$0 = \mathbf{x}^T 0 = \mathbf{x}^T A^T A\mathbf{x} = (A\mathbf{x})^T (A\mathbf{x}) = ||\, A\mathbf{x}\, ||^2$$

so that $A\mathbf{x} = 0$. However, we know by Theorem 1.4.15 that the homogeneous system with A as its coefficient matrix must have a unique solution. Of course, this solution is the zero vector. Therefore, $\mathbf{x} = 0$. It follows that the square matrix $A^T A$ has rank n (and is invertible as well) by Theorem 2.5.9 □

EXAMPLE 4.2.12. Two parameters, x_1 and x_2, are linearly related. Three samples are taken that lead to the system of equations

$$\begin{aligned} 2x_1 + x_2 &= 0 \\ x_1 + x_2 &= 0 \\ 2x_1 + x_2 &= 2 \end{aligned}$$

Show this system is inconsistent, and find the least squares solution for $\mathbf{x} = (x_1, x_2)$. What is the minimum norm of the residual $\mathbf{b} - A\mathbf{x}$ in this case?

SOLUTION. In this case it is obvious that the system is inconsistent: the first and third equations have the same quantity, $2x_1 + x_2$, equal to different values 0 and 2. Of course, we could have set up the augmented matrix of the system and found a pivot in the right hand side column as well. We see that the (rank 2) coefficient matrix A and right hand side $\mathbf{b}$ are

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

Thus

$$A^T A = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 9 & 5 \\ 5 & 3 \end{bmatrix}$$

and

$$A^T \mathbf{b} = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

As predicted by the preceding theorem, A^TA is invertible and we recall the 2×2 formula for the inverse:

$$(A^TA)^{-1} = \begin{bmatrix} 9 & 5 \\ 5 & 3 \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} 3 & -5 \\ -5 & 9 \end{bmatrix}$$

so that the unique least squares solution is

$$\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b} = \frac{1}{2}\begin{bmatrix} 3 & -5 \\ -5 & 9 \end{bmatrix}\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

The minimum value for the residual $\mathbf{b} - A\mathbf{x}$ occurs when $\mathbf{x}$ is a least squares solution, so we get

$$\begin{aligned}\mathbf{b} - A\mathbf{x} &= \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\end{aligned}$$

and therefore

$$\|\,\mathbf{b} - A\mathbf{x}\,\| = \sqrt{2} = \approx 1.414$$

This isn't terribly small, but it's the best we can do with this system. This number tells us the system is badly inconsistent. □

4.2 Exercises

1. Determine if the following pairs of vectors are orthogonal, and if so, verify that the Pythagorean theorem holds for the pair. If not, use the projection formula (which is valid even in the complex case – assume this) to find the projection of $\mathbf{u}$ along the vector $\mathbf{v}$ and express $\mathbf{u}$ as the sum of a vector parallel to $\mathbf{v}$ and a vector orthogonal to $\mathbf{v}$.

(a) $\mathbf{u} = (-2, 1, 3)$ and $\mathbf{v} = (1, 2, 0)$ (b) $\mathbf{u} = (i, 2)$ and $\mathbf{v} = (2, i)$

(c) $\mathbf{u} = (1, 1, 0, -1)$ and $\mathbf{v} = (1, -1, 3, 0)$ (d) $\mathbf{u} = (i, 1)$ and $\mathbf{v} = (1, -i)$

2. Let $\mathbf{v}_1 = [1, 0, 1]^T$ and $\mathbf{v}_2 = [1, 1, -1]^T$.

(a) Find the cosine of the angle between the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

(b) Find unit vectors in the directions of $\mathbf{v}_1$ and $\mathbf{v}_2$.

(c) Find the projection and component of the vector $\mathbf{v}_1$ along $\mathbf{v}_2$.

(d) Verify the CBS Inequality for the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

3. Repeat Exercise 2 with $\mathbf{v}_1 = (-1, 0, 2)$ and $\mathbf{v}_2 = (1, 1, -1)$.

4. For the following, find the normal equations and solve them for the system $A\mathbf{x} = \mathbf{b}$. Also find the residual vector and its norm in each case. (Note: these systems need not have a unique least squares solution.)

(a) $A = \begin{bmatrix} 2 & -2 \\ 1 & 1 \\ 3 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$ (b) $A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$

(c) $A = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ (d) $A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 2 \\ 1 & 2 & 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$

5. Show that if two vectors $\mathbf{u}$ and $\mathbf{v}$ satisfy the equation $||\mathbf{u}+\mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$, then $\mathbf{u}$ and $\mathbf{v}$ must be orthogonal. *Hint:* Express each norm in terms of dot products.

6. Suppose that you have collected data points (x_k, y_k) that are theoretically linearly related by a line of the form $y = ax + b$. Each data point gives an equation for a and b. Suppose the collected data points are $(0, .3), (1, 1.1), (2, 2), (3, 3.5)$, and $(3.5, 3.6)$. Write out the system of 5 equations that result, compute the normal equations and solve them to find the line that best fits this data. A calculator or computer might be helpful.

7. Let A be an $m \times n$ real matrix and $B = A^T A$. Show the following

(a) The matrix B is nonnegative definite.

(b) If A has full column rank, then B is positive definite.

8. Show that the CBS inequality is valid for complex vectors $\mathbf{u}$ and $\mathbf{v}$ by evaluating the nonnegative expression $||\mathbf{u} + c\mathbf{v}||^2$ with the complex dot product and evaluating it at $c = ||\mathbf{u}||^2 / (\mathbf{u} \cdot \mathbf{v})$ in the case $\mathbf{u} \cdot \mathbf{v} \neq 0$.

9. In Example 4.2.10 two values of c are calculated: The average value and the best squares value. Calculate the resulting residual and its norm in each case.

10. Show that if A is a rank 1 real matrix, then the normal equations with coefficient matrix A are consistent. *Hint:* Use Exercise 13.

11. Show that if $\mathbf{u}$ and $\mathbf{v}$ are vectors of the same length, then $\mathbf{u}+\mathbf{v}$ is orthogonal to $\mathbf{u}+\mathbf{v}$. Sketch a picture in the plane and interpret this result geometrically.

12. Verify that the projection formula (Theorem 4.2.8) is valid for complex vectors.

13. If A is a real matrix, then $A^T A$ is symmetric nonnegative definite.

14. If A is a real matrix, then $A^T A$ is positive definite if and only if A has full column rank.

4.3. Unitary and Orthogonal Matrices

Orthogonal Sets of Vectors

In our discussion of bases in Chapter 3, we saw that linear independence of a set of set of vectors was a key idea for understanding the nature of vector spaces. One of our examples of a linearly independent set (a basis, actually) was the standard basis $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ of $\mathbb{R}^n$. Here $\mathbf{e}_i$ is the vector with a 1 in the ith coordinate and zeros elsewhere. In the case of geometrical vectors and $n = 3$, these are just the familiar $\mathbf{i}, \mathbf{j}, \mathbf{k}$. These vectors have some particularly nice properties that go beyond linear independence. For one, each is a unit vector with respect to the standard norm. Furthermore, these vectors are pairwise orthogonal to each other. These properties are so desirable that we elevate them to the status of a definition.

DEFINITION 4.3.1. The set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in a standard vector space are said to be an *orthogonal set* if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ whenever $i \neq j$. If, in addition, each vector has unit length, i.e., $\mathbf{v}_i \cdot \mathbf{v}_i = 1$ then the set of vectors is said to be an *orthonormal set* of vectors.

EXAMPLE 4.3.2. Which of the following sets of vectors are orthogonal? Orthonormal? Use the standard inner product in each case.

$$\text{(a) } \{(3/5, 4/5), (-4/5, 3/5)\} \quad \text{(b) } \{(1, -1, 0), (1, 1, 0), (0, 0, 1)\}$$

SOLUTION. In the case of (a) we let $\mathbf{v}_1 = (3/5, 4/5)$, $\mathbf{v}_2 = (-4/5, 3/5)$ to obtain that

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \frac{-12}{25} + \frac{12}{25} = 0 \text{ and } \mathbf{v}_1 \cdot \mathbf{v}_1 = \frac{9}{25} + \frac{16}{25} = 1 = \mathbf{v}_2 \cdot \mathbf{v}_2.$$

It follows that the first set of vectors is an orthonormal set.

In the case of (ii) we let $\mathbf{v}_1 = (1, -1, 0), \mathbf{v}_2 = (1, 1, 0), \mathbf{v}_3 = (0, 0, 1)$ and see that

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = 1 \cdot 1 - 1 \cdot 1 + 0 \cdot 0 = 0 \text{ and } \mathbf{v}_1 \cdot \mathbf{v}_3 = 1 \cdot 0 - 1 \cdot 0 + 0 \cdot 1 = 0 = \mathbf{v}_2 \cdot \mathbf{v}_3$$

Hence this set of vectors is orthogonal, but $\mathbf{v}_1 \cdot \mathbf{v}_1 = 1 \cdot 1 + (-1) \cdot (-1) + 0 = 2$, which is sufficient to show that the vectors do not form an orthonormal set. □

One of the principal reasons that orthogonal sets are so desirable is the following key fact, which we call the *orthogonal coordinates theorem.*

Orthogonal Coordinates Theorem

THEOREM 4.3.3. *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be an orthogonal set of nonzero vectors and suppose that $\mathbf{v} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Then $\mathbf{v}$ can be expressed uniquely (up to order) as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, namely*

$$\mathbf{v} = \frac{\mathbf{v}_1 \cdot \mathbf{v}}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{v}_2 \cdot \mathbf{v}}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 + \ldots + \frac{\mathbf{v}_n \cdot \mathbf{v}}{\mathbf{v}_n \cdot \mathbf{v}_n}\mathbf{v}_n$$

PROOF. Since $v \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, we know that $\mathbf{v}$ is expressible as *some* linear combination of the v_i's, say

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n$$

Now we carry out a simple but wonderful trick that one sees used frequently with orthogonal sets, namely, take the inner product of both sides with the vector $\mathbf{v}_k$. Also, we have that $\langle \mathbf{v}_k, \mathbf{v}_j \rangle = 0$ if $j \neq k$, so we obtain

$$\begin{aligned} \mathbf{v}_k \cdot \mathbf{v} &= \mathbf{v}_k \cdot (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n) \\ &= c_1\mathbf{v}_k \cdot \mathbf{v}_1 + c_2\mathbf{v}_k \cdot \mathbf{v}_2 + \ldots + c_n\mathbf{v}_k \cdot \mathbf{v}_n \\ &= c_k\mathbf{v}_k \cdot \mathbf{v}_k \end{aligned}$$

Since $\mathbf{v}_k \neq 0$, it follows that $\mathbf{v}_k \cdot \mathbf{v}_k \neq 0$, so that we may solve for c_k to obtain that

$$c_k = \frac{\mathbf{v}_k \cdot \mathbf{v}}{\mathbf{v}_k \cdot \mathbf{v}_k}$$

This proves that the coefficients c_k are unique and establishes the formula of the theorem. □

COROLLARY 4.3.4. *Every orthogonal set of nonzero vectors is linearly independent.*

PROOF. Consider a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. If some linear combination were to sum to zero, say

$$0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n$$

it would follow from the preceding theorem that

$$c_k = \frac{\mathbf{v}_k \cdot 0}{\mathbf{v}_k \cdot \mathbf{v}_k} = 0$$

It follows from the definition of linear independence that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent. □

Several observations are worth noting:

- The converse of the corollary is false, that is, not every linearly independent set of vectors is orthogonal. For an example, consider the linearly independent vectors $\mathbf{v}_1 = (1, 0)$, $\mathbf{v}_2 = (1, 1)$ in $V = \mathbb{R}^2$.
- The vector $\frac{\mathbf{v}_k \cdot \mathbf{v}}{\mathbf{v}_k \cdot \mathbf{v}_k}\mathbf{v}_k$ looks familiar. In fact, it is the projection of the vector $\mathbf{v}$ along the vector $\mathbf{v}_k$. Thus, we can say Theorem 6.2.18 in words as follows: any linear combination of an orthogonal set of nonzero vectors is the sum of its projections in the direction of each vector in the set.
- The coefficients c_k of Theorem 6.2.18 are also familiar: they are the *coordinates* of v relative to the basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. This terminology was introduced in Section 3.3 of Chapter 3. Thus Theorem 6.2.18 shows us that coordinates are rather easy to calculate with respect to an orthogonal basis. Contrast this with Example 3.3.13 of Chapter 3.
- The formula of Theorem 6.2.18 simplifies very nicely if vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ form an orthonormal set (which automatically consists of nonzero vectors!), namely

$$\mathbf{v} = \mathbf{v}_1 \cdot \mathbf{v}\,\mathbf{v}_1 + \mathbf{v}_2 \cdot \mathbf{v}\,\mathbf{v}_2 + \ldots + \mathbf{v}_n \cdot \mathbf{v}\,\mathbf{v}_n$$

- Given an orthogonal set of nonzero vectors, it is easy to manufacture an orthonormal set of vectors. Simply replace every vector in the original set by the vector divided by its length.

Orthogonal and Unitary Matrices

In general, if we want to determine the coordinates of a vector $\mathbf{b}$ with respect to a certain basis of vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$, we stack the basis vectors together to form a matrix A, then solve the system $A\mathbf{x} = \mathbf{b}$ for the vector of coordinates $\mathbf{x}$ of $\mathbf{b}$ with respect to this basis. In fact, $\mathbf{x} = A^{-1}\mathbf{b}$. Now we have seen that if the basis vectors happen to form an orthonormal set, the situation is much simpler and we certainly don't have to find A^{-1}. Is this simplicity reflected in properties of the matrix A? The answer is "yes" and we can see this as follows: suppose that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ is an orthonormal basis of $\mathbb{R}^n$ and let $A = [\,\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$. Orthogonality says that $\mathbf{u}_i^T\mathbf{u}_j = \delta_{ij}$. This means that the matrix $A^T A$, whose (i,j)th entry is $\mathbf{u}_n^T\mathbf{u}_n$, is simply $[\delta_{ij}] = I$, that is, $A^T A = I$. Now recall that Theorem 2.5.9 of Chapter 2 shows that if a matrix acts as an inverse on one side and the matrices in question are square, then the matrix really is the two-sided inverse. Hence, $A^{-1} = A^T$. A similar argument works if $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ is an orthonormal basis of $\mathbb{C}^n$ and we use Hermitian transpose instead of transpose. Matrices bearing these properties are important enough to have their own names.

DEFINITION 4.3.5. A square matrix U is called *unitary* if $U^H = U^{-1}$ and Q is called *orthogonal* if Q is real and $Q^T = Q^{-1}$.

One could allow orthogonal matrices to be complex as well, but these are not particularly useful for us, so in this text we will always assume that orthogonal matrices have real entries. Since for real matrices Q, we have $Q^H = Q^T$, we see from the definition that orthogonal matrices are exactly the real unitary matrices.

The naming is traditional in matrix theory, but a bit unfortunate because it sometimes causes confusion between the terms "orthogonal vectors" and "orthogonal matrix." By orthogonal *vectors* we mean a set of vectors with a certain relationship to each other, while an orthogonal *matrix* is a real matrix whose inverse is its transpose. And to make matters more confusing, there is a close connection between the two terms, since a square matrix is orthogonal exactly when its columns form an orthonormal set.

EXAMPLE 4.3.6. Show that the matrix $U = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$ is unitary and that for any angle θ, the matrix $R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal.

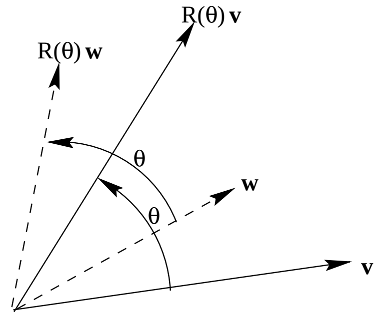


FIGURE 4.3.1. Action of rotation matrix $R(\theta)$.

SOLUTION. It is sufficient to check that $U^H U = I$ and $R(\theta)^T R(\theta) = I$. So we calculate

$$\begin{aligned} U^H U &= \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \right)^H \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -i \\ -i & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 1 - i^2 & i - i \\ -i + i & 1 - i^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

which shows that U is unitary. For the real matrix $R(\theta)$ we have

$$\begin{aligned} R(\theta)^T R(\theta) &= \left(\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \right)^T \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos^2\theta + \sin^2\theta & \cos\theta\sin\theta - \sin\theta\cos\theta \\ -\cos\theta\sin\theta + \sin\theta\cos\theta & \cos^2\theta + \sin^2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \end{aligned}$$

which shows that $R(\theta)$ is orthogonal. □

Orthogonal and unitary matrices have a certain “rigidity” quality about them which is nicely illustrated by the rotation matrix $R(\theta)$ that turns up in calculus as coefficients for a rotational change of basis. The effect of multiplying a vector $\mathbf{x} \in \mathbb{R}^2$ by $R(\theta)$ is to rotate the vector counterclockwise through an angle of θ as illustrated in Figure 4.3.1. In particular, angles between vectors and lengths of vectors are preserved by such a multiplication. This is no accident of $R(\theta)$, but rather a property of orthogonal and unitary matrices in general. Here is a statement of these properties for orthogonal matrices. An analogous fact holds for complex unitary matrices with vectors in $\mathbb{C}^n$.

THEOREM 4.3.7. *Let Q be an orthogonal $n \times n$ matrix and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ with the standard inner (dot) product. Then*

$$||Q\mathbf{x}|| = ||\mathbf{x}|| \qquad \textit{and} \qquad Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$$

PROOF. Let us calculate the norm of $Q\mathbf{x}$:

$$||Q\mathbf{x}||^2 = Q\mathbf{x}\cdot Q\mathbf{x} = (Q\mathbf{x})^T\, Q\mathbf{x} = \mathbf{x}^T Q^T Q\mathbf{x} = \mathbf{x}^T\mathbf{x} = ||\mathbf{x}||^2$$

which proves the first assertion, while similarly

$$Q\mathbf{x}\cdot Q\mathbf{y} = (Q\mathbf{x})^T\, Q\mathbf{y} = \mathbf{x}^T Q^T Q\mathbf{y} = \mathbf{x}^T\mathbf{y} = \mathbf{x}\cdot\mathbf{y}$$

□

There is one more kind of orthogonal matrix that has turned out to be very useful in numerical calculations and has a very nice geometrical interpretation as well. It gives us a very simple way of forming orthogonal matrices directly.

DEFINITION 4.3.8. A matrix of the form $H_{\mathbf{v}} = I - 2(\mathbf{v}\mathbf{v}^T)/(\mathbf{v}^T\mathbf{v})$, where $\mathbf{v}\in\mathbb{R}^n$, is called a *Householder* matrix.

EXAMPLE 4.3.9. Let $\mathbf{v} = (3, 0, 4)$ and compute the Householder matrix $H_{\mathbf{v}}$. What is the effect of multiplying it by the vector $\mathbf{v}$?

SOLUTION. We calculate $H_{\mathbf{v}}$ to be

$$\begin{aligned} I - \frac{2}{\mathbf{v}^T\mathbf{v}}\mathbf{v}\mathbf{v}^T &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{2}{3^2+4^2}\begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix}\begin{bmatrix} 3 & 0 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{2}{25}\begin{bmatrix} 9 & 0 & 12 \\ 0 & 0 & 0 \\ 12 & 0 & 16 \end{bmatrix} \\ &= \frac{1}{25}\begin{bmatrix} 7 & 0 & -24 \\ 0 & 25 & 0 \\ -24 & 0 & -7 \end{bmatrix} \end{aligned}$$

Thus we have that multiplying $H_{\mathbf{v}}$ by $\mathbf{v}$ gives

$$H_{\mathbf{v}}\mathbf{v} = \frac{1}{25}\begin{bmatrix} 7 & 0 & -24 \\ 0 & 25 & 0 \\ -24 & 0 & -7 \end{bmatrix}\begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix} = \frac{1}{25}\begin{bmatrix} -75 \\ 0 \\ -100 \end{bmatrix} = -\begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix}$$

□

The behavior of this example is no accident. Multiplication by a Householder matrix can be thought of as a geometrical reflection that reflects the vector $\mathbf{v}$ to $-\mathbf{v}$ and leaves any vector orthogonal to $\mathbf{v}$ unchanged. This is implied by the following theorem. For a picture of this geometrical interpretation, see Figure 4.3.2. Notice that in this figure V is the plane perpendicular to $\mathbf{v}$ and the reflections are across this plane.

THEOREM 4.3.10. *Let $H_{\mathbf{v}}$ be the Householder matrix defined by $\mathbf{v}\in\mathbb{R}^n$ and let $\mathbf{w}\in\mathbb{R}^n$ be written as $\mathbf{w} = \mathbf{p}+\mathbf{u}$, where $\mathbf{p}$ is the projection of $\mathbf{w}$ along $\mathbf{v}$ and $\mathbf{u} = \mathbf{w}-\mathbf{p}$. Then*

$$H_{\mathbf{v}}\mathbf{w} = -\mathbf{p}+\mathbf{u}$$

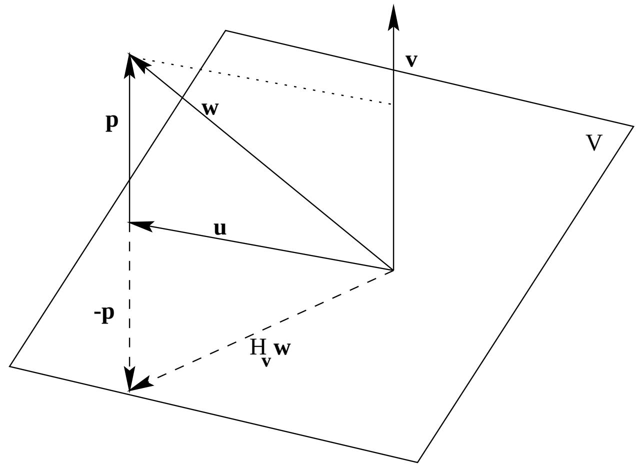


FIGURE 4.3.2. Action of $\mathbf{H}_{\mathbf{v}}$ on $\mathbf{w}$ as a reflection across the plane V perpendicular to $\mathbf{v}$.

PROOF. With notation as in the statement of the theorem, we have $\mathbf{p} = \frac{\mathbf{v}^T\mathbf{w}}{\mathbf{v}^T\mathbf{v}}\mathbf{v}$ and $\mathbf{w} = \mathbf{p} + \mathbf{u}$. Let us calculate

$$
\begin{aligned}
H_{\mathbf{v}}\mathbf{w} &= (I - \frac{2}{\mathbf{v}^T\mathbf{v}}\mathbf{v}\mathbf{v}^T)(\mathbf{p} + \mathbf{u}) \\
&= \mathbf{p} + \mathbf{u} - 2\frac{\mathbf{v}^T\mathbf{w}}{(\mathbf{v}^T\mathbf{v})^2}\mathbf{v}\mathbf{v}^T\mathbf{v} - 2\frac{\mathbf{v}^T\mathbf{w}}{\mathbf{v}^T\mathbf{v}}\mathbf{v}\mathbf{v}^T\mathbf{u} \\
&= \mathbf{p} + \mathbf{u} - 2\frac{\mathbf{v}^T\mathbf{w}}{\mathbf{v}^T\mathbf{v}}\mathbf{v} - \mathbf{0} \\
&= \mathbf{p} + \mathbf{u} - 2\mathbf{p} \\
&= \mathbf{u} - \mathbf{p}
\end{aligned}
$$

□

EXAMPLE 4.3.11. Let $\mathbf{v} = (3, 0, 4)$ and $H_{\mathbf{v}}$ the corresponding Householder matrix (as in Example 4.3.9). The columns of this matrix form an orthonormal basis for the space $\mathbb{R}^3$. Find the coordinates of the vector $\mathbf{w} = (2, 1, -4)$ relative to this basis.

SOLUTION. We have already calculated $H_{\mathbf{v}} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ in Example 4.3.9. The vector $\mathbf{c} = (c_1, c_2, c_3)$ of coordinates of $\mathbf{w}$ must satisfy the equations

$$\mathbf{w} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 = H_{\mathbf{v}}\mathbf{c}.$$

Since $H_{\mathbf{v}}$ is orthogonal, it follows that

$$\mathbf{c} = H_{\mathbf{v}}^{-1}\mathbf{w} = H_{\mathbf{v}}^T = \frac{1}{25}\begin{bmatrix} 7 & 0 & -24 \\ 0 & 25 & 0 \\ -24 & 0 & -7 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix} = \begin{bmatrix} 4.56 \\ 1 \\ -0.8 \end{bmatrix}$$

which gives us the required coordinates. □

For the most part we work with real Householder matrices. However, occasionally complex numbers are a necessary part of the scenery. In such situations we can define the *complex* Householder matrix by the formula

$$H_{\mathbf{v}} = I - 2(\mathbf{v}\mathbf{v}^H)/(\mathbf{v}^H\mathbf{v})$$

The projection formula (Theorem 4.2.8) remains valid for complex vectors which is all we need to see that the proof of Theorem 4.3.10 carries over to complex vectors provided that we replace all transposes by Hermitian transposes.

One might ask if there is any other way to generate orthogonal matrices. In particular, if we start with a single unit vector, can we embed it as a column in an orthogonal matrix? The answer is "yes," and truth of this answer follows from an even stronger statement, which is reminiscent of the Steinitz substitution principle.

THEOREM 4.3.12. *Every orthogonal set of nonzero vectors in a standard vector space can be expanded to an orthogonal basis of the space.*

PROOF. Suppose that the space in question is $\mathbb{R}^n$ and we have expanded our original orthogonal set to $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, where $k < n$. We show how to add one more element. This is sufficient, because by repeating this step we eventually fill up $\mathbb{R}^n$. Let $A = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]^T$ and let $\mathbf{v}_{k+1}$ be any nonzero solution to $A\mathbf{x} = 0$, which exists since $k < n$. This vector is orthogonal to the $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$. □

We'll see a more efficient way to perform this calculation when we study the Gram-Schmidt algorithm in Chapter 6.

EXAMPLE 4.3.13. The vectors $\mathbf{u}_1 = \frac{1}{3}(1,2,2)$ and $\mathbf{u}_2 = \frac{1}{\sqrt{5}}(-2,1,0)$ form an orthonormal set. Find an orthogonal matrix with these vectors as the first two columns.

SOLUTION. To keep the arithmetic simple, let $\mathbf{v}_1 = (1,2,2)$ and $\mathbf{v}_2 = (-2,1,0)$. Form the matrix A with these vectors as rows and solve the system $A\mathbf{x} = 0$ to get a general solution (the reader should check this) $\mathbf{x} = (-\frac{2}{5}x_3, -\frac{4}{5}x_3, x_3)$. So take $x_3 = 5$ and get a particular solution $\mathbf{v}_3 = (-2,-4,5)$. Now normalize all three vectors $\mathbf{v}_j$ to recover the original $\mathbf{u}_1, \mathbf{u}_2$ and the new $\mathbf{u}_3 = \frac{1}{3\sqrt{5}}(-2,-4,5)$. Stack these columns together and factor out $\frac{1}{3\sqrt{5}}$ to obtain the orthogonal matrix

$$P = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3] = \frac{1}{3\sqrt{5}} \begin{bmatrix} \sqrt{5} & -6 & -2 \\ 2\sqrt{5} & 3 & -4 \\ 2\sqrt{5} & 0 & 5 \end{bmatrix}$$

which is the matrix we want. □

4.3 Exercises

1. Which of the following sets of vectors are linearly independent? Orthogonal? Orthonormal?

(a) $(1,-1,2), (2,2,0)$ (b) $(3,-1,1), (1,2,-1), (2,-1,0)$

(c) $\frac{1}{5}(3,4), \frac{1}{5}(4,-3)$ (d) $(1+i,1), (1,1-i)$

2. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix}$. Show this set is an orthogonal basis of $\mathbb{R}^3$ and find the coordinates of $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$ with respect to this basis.

3. For what values of the angle θ is the orthogonal matrix $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ symmetric? Skew-symmetric?

4. Determine which of the following matrices are orthogonal or unitary. For such matrices, find their inverses.

(a) $\frac{1}{5}\begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix}$ (b) $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$ (c) $\frac{1}{\sqrt{3}}\begin{bmatrix} 1+i & i \\ i & 1-i \end{bmatrix}$

(d) $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2}i & 0 \\ i & 0 & -i \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix}$ (f) $\frac{1}{2}\begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}$

5. Let $\mathbf{u} = (1, 2, -1)$ and $\mathbf{w} = (\sqrt{6}, 0, 0)$. Let $\mathbf{v} = \mathbf{u} - \mathbf{w}$ and construct the Householder matrix $H_{\mathbf{v}}$. Now apply it to the vectors $\mathbf{u}$ and $\mathbf{w}$. Conclusions?

6. Find orthogonal or unitary matrices that include the following orthonormal vectors in their columns

(a) $\mathbf{u}_1 = \frac{1}{\sqrt{6}}(1, 2, -1)$, $\mathbf{u}_2 = \frac{1}{\sqrt{3}}(-1, 1, 1)$ (b) $\mathbf{u}_1 = \frac{1}{5}(3, -4)$ (c) $\mathbf{u}_1 = \frac{1}{2}(1 + i, 1 - i)$

(d) $\mathbf{u}_1 = \frac{1}{\sqrt{3}}(1, 1, 1)$, (e) $\mathbf{u}_1 = \frac{1}{2}(1, 1, -1, -1)$, $\mathbf{u}_2 = \frac{1}{2}(1, -1, 1, -1)$, $\mathbf{u}_3 = \frac{1}{\sqrt{2}}(0, 1, 1, 0)$

7. Show that if P is an orthogonal matrix, then $e^{i\theta}P$ is a unitary matrix for any real θ.

8. Let $P = \frac{1}{2}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$. Verify that P is a *projection matrix* , that is, $P^T = P$ and $P^2 = P$, and that if $R = I - 2P$, then R is a *reflection matrix*, that is, R is a symmetric orthogonal matrix.

9. Let P be a real projection matrix and $R = I - 2P$. Prove that R is a reflection matrix. (See Exercise 8 for definitions.)

10. Let $R = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ and $P = \frac{1}{2}(I - R)$. Verify that R is a reflection matrix and P is a projection matrix. (See Exercise 8 for definitions.)

11. Let R be a reflection matrix. Prove that $P = \frac{1}{2}(I - R)$ is a projection matrix.

12. Prove that every Householder matrix is a reflection matrix.

13. Let U and V be orthogonal matrices.

(a) Show that the product UV is also orthogonal.

(b) Find examples of orthogonal matrices U and V whose sum is not an orthogonal matrix.

14. Let the quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ be defined by the formula $y = f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, where A is a real matrix. Suppose that an orthogonal change of variables is made from in both domain and range, say $x = Qx'$ and $\mathbf{y} = Q\mathbf{y}'$, where Q is orthogonal. Show that in the new coordinates $\mathbf{x}'$ and $\mathbf{y}'$, $\mathbf{y}' = \mathbf{x}'^T(Q^T AQ)\mathbf{x}'$

4.4. *Computational Notes and Projects

Project: Least Squares

The Big Eight needs your help! Below is a table of scores from the games played thus far: The (i, j)th entry is team i's score in the game with team j. Your assignment is two-fold. First, write a notebook that contains instructions for the illiterate on how to plug in known data and obtain team ratings and predicted point spreads based on the least squares and graph theory ideas you have seen. Secondly, you are to write a brief report (one to three pages) on your project which describes the problem, your solution to it, its limitations and the ideas behind it.

	CU	IS	KS	KU	MU	NU	OS	OU
CU		24		21	45		21	14
IS	12			42	21	16		7
KS				12	21	3	27	24
KU	9	14	30			10		14
MU	8	3	52			18	21	
NU		51	48	63	26		63	
OS	41		45		49	42		28
OU	17	35	70	63			31	

Implementation Notes: You will need to set up suitable system of equations, form the normal equations, and have a computer algebra system solve the problem. For purposes of illustration, we assume in this project that the tool in use is Mathematica. If not, you will need to replace these commands with the appropriate ones that your computational tools provide. The equations in question are formed by letting the variables be a vector $\mathbf{x}$ of "potentials" $x(i)$, one for each team i, so that the "potential differences" best approximate the actual score differences (i.e., point spreads) of the games. To find the vector $\mathbf{x}$ of potentials you solve the system $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b}$ is the vector of observed potential differences. N.B: the matrix A is *not* the table given above. You will get one equation for each game played. For example, by checking the $(1, 2)$th and $(2, 1)$th entries, we see that CU beat IS by a score of 24 to 12. So the resulting equation for this game is $x(1) - x(2) = 24 - 12 = 12$. Ideally, the resulting potentials would give you numbers that would enable you to predict the point spread of an as yet unplayed game: all you would have to do to determine the spread for team i versus team j is calculate the difference $x(j) - x(i)$. Of course, it doesn't really work out this way, but this is a good use of the known data. When you set up this system, you obtain an inconsistent system. This is where least squares enter the picture. You will need to set up and solve the normal equations, one way or another. You might notice that the null space of the resulting coefficient matrix is nontrivial, so this matrix is not full column rank. This makes sense: potentials are unique up to a constant. To fix this, you could arbitrarily fix the value of one team's potential. E.g., set the weakest team's potential value to zero by adding one additional equation to the system of the form $x(i) = 0$.

Notes to the Instructor: the data above came from the now defunct Big Eight Conference. This project works better when adapted to your local environment. Pick a sport in season at your institution or locale. Have students collect the data themselves, make out a data table as above, and predict the spread for some (as yet) unplayed games of local interest. It can be very interesting to make it an ongoing project, where for a number of weeks the students are required to collect last week's data and make predictions for next week based on all data collected to date.

4.4 Exercises

1. It is hypothesized that sale of a certain product is linearly dependent on three factors. The sales output is quantified as z and the three factors as x_1, x_2 and x_3. Six samples are taken of the sales and the factor data. Results are contained in the following table. Does the linearity hypothesis seem reasonable? Explain your answer.

z	x_1	x_2	x_3
527	13	5	6
711	6	17	7
1291	12	16	23
625	11	13	4
1301	12	27	14
1350	5	14	31

Review

Chapter 4 Exercises

1. Find a unit vector orthogonal and a unit vector parallel to $[1, 3, 2]^T$ in $\mathbb{R}^3$.

2. Let $\mathbf{u} = [1, 2, -1, 1]^T$ and $\mathbf{v} = [-2, 1, 0, 0]^T$ and compute $||\mathbf{u}||$, $\mathbf{u} \cdot \mathbf{v}$, and the angle between these vectors.

3. Find the projection of $\mathbf{u} = (1, 2, 0, 1)$ along $\mathbf{v} = (1, 1, 1, 1)$ and the projection of $\mathbf{v}$ along $\mathbf{u}$ and express $\mathbf{v}$ as the sum of a vector parallel to $\mathbf{u}$ and a vector orthogonal to $\mathbf{u}$.

4. Let $\mathbf{u}, \mathbf{v}$ be linearly independent vectors in a standard vector space and let $W = \mathrm{span}\{\mathbf{u}, \mathbf{v}\}$. Show that $\mathbf{u}, \mathbf{v} - ((\mathbf{u} \cdot \mathbf{v})/(\mathbf{v} \cdot \mathbf{v}))\mathbf{u}$ is an orthogonal basis of W.

5. Let $W = \mathrm{span}\{(1, 2, 1), (2, -1, 0)\}$

(a) Show this spanning set is an orthogonal set.

(b) The vector $\mathbf{v} = (-4, 7, 2)$ belongs to W. Calculate its coordinates with respect to this basis.

6. Determine if $W = \{\mathbf{v} \in \mathbb{R}^3 \mid ||\mathbf{v}|| = 1\}$ is a subspace of $\mathbb{R}^3$.

7. Find an orthogonal matrix which has as its first column the vector $\frac{1}{3}(1, 0, 2, -2)$.

8. Show that if A is a real symmetric $n \times n$ matrix and $\mathbf{u}, \mathbf{v}$ are vectors in $\mathbb{R}^n$, then

$$(A\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot A\mathbf{v}$$

CHAPTER 5

THE EIGENVALUE PROBLEM

The first major problem of linear algebra is to understand how to solve the basis linear system $A\mathbf{x} = \mathbf{b}$ and what the solution means. We have explored this system from three points of view: in Chapter 1 we approached the problem from an operational point of view and learned the mechanics of computing solutions. In Chapter 2, we took a more sophisticated look at the system from the perspective of matrix theory. Finally, in Chapter 3, we viewed the problem from the vantage of vector space theory.

Now it time for us to begin study of the second major problem of linear algebra, namely the eigenvalue problem. It was necessary for us to tackle the linear systems problem first because the eigenvalue problem is more sophisticated and will require most of the tools that we have thus far developed. This subject has many useful applications; indeed, it arose out of these applications. One of the more interesting applications of eigenvalue theory that we study in this chapter is the analysis of discrete dynamical systems. Such systems include the Markov chains we have seen in earlier chapters as a special case.

5.1. Definitions and Basic Properties

What are They?

Good question. Let's get right to the point.

Eigenvalues and Eigenvectors

DEFINITION 5.1.1. Let A be a square $n \times n$ matrix. An *eigenvector* of A is a nonzero vector $\mathbf{x}$ in $\mathbb{R}^n$ (or $\mathbb{C}^n$, if we are working over complex numbers), such that for some scalar λ , we have

$$A\mathbf{x} = \lambda\mathbf{x}$$

The scalar λ is called an *eigenvalue* of the matrix A, and we say that the vector $\mathbf{x}$ is an *eigenvector belonging to the eigenvalue* λ. The pair $(\lambda, \mathbf{x})$ is called an *eigenpair* for the matrix A.

The only kinds of matrices for which these objects are defined are square matrices, so we'll assume throughout this Chapter that we are dealing with such matrices.

Caution: Be aware that the eigenvalue λ is allowed to be the 0 scalar, but eigenvectors $\mathbf{x}$ are, by definition, *never the* 0 *vector*.

As a matter of fact, it is quite informative to have an eigenvalue 0. This says that the system $A\mathbf{x} = 0\mathbf{x} = 0$ has a nontrivial solution $\mathbf{x}$, in other words, A is not invertible by Theorem 2.5.9.

Here are a few simple examples of eigenvalues and vectors. Let $A = \begin{bmatrix} 7 & 4 \\ 3 & 6 \end{bmatrix}$, $\mathbf{x} = (-1, 1)$ and $\mathbf{y} = (4, 3)$. One checks that $A\mathbf{x} = (-3, 3) = 3\mathbf{x}$ and $A\mathbf{y} = (40, 30) = 10\mathbf{y}$. It follows that $\mathbf{x}$ and $\mathbf{y}$ are eigenvectors corresponding to eigenvalues 3 and 10, respectively.

Why should we have any interest in these quantities? A general answer goes something like this: knowledge of eigenvectors and eigenvalues gives us deep insights into the structure of the matrix A. Here is just one example: suppose that we would like to have a better understanding of the effect of multiplication of a vector $\mathbf{x}$ by powers of the matrix A, that is, of $A^k\mathbf{x}$. Let's start with the first power, $A\mathbf{x}$. If we knew that $\mathbf{x}$ were an eigenvector of A, then we would have that for some scalar λ,

$$\begin{aligned} A\mathbf{x} &= \lambda\mathbf{x} \\ A^2\mathbf{x} &= A(A\mathbf{x}) = A\lambda\mathbf{x} = \lambda A\mathbf{x} = \lambda^2\mathbf{x} \\ &\vdots \\ A^k\mathbf{x} &= A(A^{k-1}\mathbf{x}) = \cdots = \lambda^k\mathbf{x} \end{aligned}$$

This is very nice, because it reduces something complicated, namely matrix-vector multiplication, to something simple, namely scalar-vector multiplication.

There are other reasons for the usefulness of the eigenvector/value concept which we will develop later, but here is one that is fairly immediate: is there any significance in knowing that one of the eigenvalues of A is 0? Check the definition of eigenvalue and we see that this means that $A\mathbf{x} = 0$ for some nonzero vector $\mathbf{x}$. By Theorem 2.5.9 of Chapter 2 (page 91) it follows that A is not invertible. So eigenvalues can tell us about invertibility.

We need some handles on these quantities. Let's ask how we could figure out what they are for specific matrices. Here are some of the basic points about eigenvalues and eigenvectors.

THEOREM 5.1.2. *Let A be a square $n \times n$ matrix. Then*

1. *The eigenvalues of A consist of all scalars λ that are solutions to the nth degree polynomial equation*
$$\det(\lambda I - A) = 0$$
2. *For a given eigenvalue λ, the eigenvectors of the matrix A belonging to that eigenvalue consist of all nonzero elements of $\mathcal{N}(\lambda I - A)$.*

PROOF. Note that $\lambda\mathbf{x} = \lambda I\mathbf{x}$. Thus we have the following chain of thought: A has eigenvalue λ if and only if $A\mathbf{x} = \lambda\mathbf{x}$, for some nonzero vector $\mathbf{x}$, which is true if and only if

$$0 = \lambda\mathbf{x} - A\mathbf{x} = \lambda I\mathbf{x} - A\mathbf{x} = (\lambda I - A)\mathbf{x}$$

for some nonzero vector $\mathbf{x}$. This last statement is equivalent to the assertion that $0 \neq \mathbf{x} \in \mathcal{N}(\lambda I - A)$. The matrix $\lambda I - A$ is square, so it has a nontrivial null space precisely when it is singular (recall the characterizations of nonsingular matrices in Theorem 2.5.9 of Chapter 2). This occurs only when $\det(\lambda I - A) = 0$. If we expand this determinant down the first column, we see that the highest order term involving λ that occurs is the product of the diagonal terms $(\lambda - a_{ii})$, so that the degree of the expression $\det(\lambda I - A)$ as a polynomial in λ is n. This proves (1).

We saw from this chain of thought that if λ is an eigenvalue of A, then the eigenvectors belonging to that eigenvalue are precisely the nonzero vectors $\mathbf{x}$ such that $(\lambda I - A)\mathbf{x} = 0$, that is, the nonzero elements of $\mathcal{N}(A)$, which is what (2) asserts. □

Here is some terminology that we will use throughout this chapter.

NOTATION 5.1.3. We call a polynomial *monic* if the leading coefficient is 1.

For instance, $\lambda^2 + 2\lambda + 3$ is a monic polynomial in λ while $2\lambda^2 + \lambda + 1$ is not.

DEFINITION 5.1.4. Given a square $n \times n$ matrix A, the equation $\det(\lambda I - A) = 0$ is called the *characteristic equation* of A and the nth degree monic polynomial $p(\lambda) = \det(\lambda I - A)$ is called the *characteristic polynomial* of A.

Suppose we already know the eigenvalues of A and want to find the eigenvalues of something like $3A + 4I$. Do we have to start over to find them? The next calculation is really a useful tool for answering such questions.

THEOREM 5.1.5. *If $B = cA + dI$ for scalars d and $c \neq 0$, then the eigenvalues of B are of the form $\mu = c\lambda + d$, where λ runs over the eigenvalues of A, and the eigenvectors of A and B are identical.*

PROOF. Let $\mathbf{x}$ be an eigenvector of A corresponding to the eigenvalue λ. Then by definition $\mathbf{x} \neq 0$ and

$$A\mathbf{x} = \lambda\mathbf{x}$$

Also, we have that

$$dI\mathbf{x} = d\mathbf{x}$$

Now multiply the first equation by the scalar c and add these two equations to obtain

$$(cA + dI)\mathbf{x} = B\mathbf{x} = (c\lambda + d)\mathbf{x}$$

It follows that every eigenvector of A belonging to λ is also an eigenvector of B belonging to the eigenvalue $c\lambda + d$. Conversely, if $\mathbf{y}$ is an eigenvalue of B belonging to μ, then

$$B\mathbf{y} = \mu\mathbf{y} = (cA + dI)\mathbf{y}$$

Now solve for $A\mathbf{y}$ to obtain that

$$A\mathbf{y} = \frac{1}{c}(\mu - d)\mathbf{y}$$

so that $\lambda = (\mu - d)/c$ is an eigenvalue of A with corresponding eigenvector $\mathbf{y}$. It follows that A and B have the same eigenvectors and their eigenvalues are related by the formula $\mu = c\lambda + d$. □

EXAMPLE 5.1.6. Let $A = \begin{bmatrix} 7 & 4 \\ 3 & 6 \end{bmatrix}$, $\mathbf{x} = (-1, 1)$ and $\mathbf{y} = (4, 3)$, so that $A\mathbf{x} = (-3, 3) = 3\mathbf{x}$ and $A\mathbf{y} = (40, 30) = 10\mathbf{y}$. Find the eigenvalues and corresponding eigenvectors for the matrix $B = 3A + 4I$.

SOLUTION. From the calculations given to us, we observe that $\mathbf{x}$ and $\mathbf{y}$ are eigenvectors corresponding to the eigenvalues 3 and 10, respectively, for A. These are all the eigenvalues of A, since the characteristic polynomial of A is of degree 2, so has only two roots. According to Theorem 5.1.5, the eigenvalues of $3A + 4I$ must be $\mu_1 = 3 \cdot 3 + 4 = 13$ with corresponding eigenvector $\mathbf{x} = (-1, 1)$, and $\mu_2 = 3 \cdot 10 + 4 = 34$ with corresponding eigenvalue $\mathbf{y} = (4, 3)$.

DEFINITION 5.1.7. Given an eigenvalue λ of the matrix A, the *eigenspace* corresponding to λ is the subspace $\mathcal{N}(\lambda I - A)$ of $\mathbb{R}^n$ (or $\mathbb{C}^n$).

Notation: We write $\mathcal{E}_\lambda(A) = \mathcal{N}(\lambda I - A)$.

DEFINITION 5.1.8. By an *eigensystem* of the matrix A, we mean a list of all the eigenvalues of A and, for each eigenvalue λ, a complete description of the eigenspace corresponding to λ.

The usual way to give a complete description of an eigenspace is to list a basis of the space. Remember that there is one element of the eigenspace $\mathcal{N}(\lambda I - A)$ that is *not* an eigenvector, namely 0. In any case, the computational route is now clear. To call it an algorithm is really an abuse of language, since we don't have a complete computational description of the root finding phase, but here it is:

Eigensystem Algorithm. Let A be an $n \times n$ matrix. To find an eigensystem of A:

1. Find the scalars that are roots to the characteristic equation $\det(\lambda I - A) = 0$.
2. For each scalar λ in (1), use the null space algorithm to find a basis of the eigenspace $\mathcal{N}(\lambda I - A)$.

As a matter of convenience, it is sometimes a little easier to work with $A - \lambda I$ when calculating eigenspaces (because there are fewer extra minus signs to worry about). This is perfectly OK, since $\mathcal{N}(A - \lambda I) = \mathcal{N}(\lambda I - A)$. It doesn't affect the eigenvalues either, since $\det(\lambda I - A) = \pm \det(A - \lambda I)$. Here is our first eigensystem calculation.

EXAMPLE 5.1.9. Find an eigensystem for the matrix $A = \begin{bmatrix} 7 & 4 \\ 3 & 6 \end{bmatrix}$.

SOLUTION. First solve the characteristic equation

$$\begin{aligned} 0 = \det(\lambda I - A) &= \det \begin{bmatrix} \lambda - 7 & -4 \\ -3 & \lambda - 6 \end{bmatrix} \\ &= (\lambda - 7)(\lambda - 6) - (-3)(-4) \\ &= \lambda^2 - 13\lambda + 42 - 12 \\ &= \lambda^2 - 13\lambda + 30 \\ &= (\lambda - 3)(\lambda - 10) \end{aligned}$$

Hence the eigenvalues are $\lambda = 3, 10$. Next, for each eigenvector calculate the corresponding eigenspace.

$\lambda = 3$: Then $A - 3I = \begin{bmatrix} 7-3 & 4 \\ 3 & 6-3 \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 3 & 3 \end{bmatrix}$ and row reduction gives

$$\begin{bmatrix} 4 & 4 \\ 3 & 3 \end{bmatrix} \xrightarrow[E_1(1/4)]{E_{21}(-3/4)} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

so the general solution is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Therefore a basis of $\mathcal{E}_3(A)$ is $\{(-1, 1)\}$.

$\lambda = 10$: Then $A - 10I = \begin{bmatrix} 7-10 & 4 \\ 3 & 6-10 \end{bmatrix} = \begin{bmatrix} -3 & 4 \\ 3 & -4 \end{bmatrix}$ and row reduction gives

$$\begin{bmatrix} -3 & 4 \\ 3 & -4 \end{bmatrix} \xrightarrow[E_1(-1/3)]{E_{21}(1)} \begin{bmatrix} 1 & -4/3 \\ 0 & 0 \end{bmatrix}$$

so the general solution is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (4/3)x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 4/3 \\ 1 \end{bmatrix}$$

Therefore a basis of $\mathcal{E}_{10}(A)$ is $\{(4/3, 1)\}$. □

Concerning this example, there are several interesting points worth noting:

1. Since the 2×2 matrix $A - \lambda I$ is singular for eigenvalue λ, one row should always be a multiple of the other. Knowing this, we didn't have to do even the little row reduction we did above. However, its a good idea to check this; it helps you avoid mistakes. Remember: any time that row reduction of $A - \lambda I$ leads to full rank (only trivial solutions) you have either made an arithmetic error or you do not have an eigenvalue.
2. This matrix is familiar. In fact, $B = (0.1)A$ is the Markov chain transition matrix from Example 2.3.4 of Chapter 2. Therefore the eigenvalues of B are 0.3 and 1, by Example 5.1.9 with $c = 0.1$ and $d = 0$. The eigenvector belonging to $\lambda = 1$ is just a solution to the equation $B\mathbf{x} = \mathbf{x}$, which was discussed in Example 3.4.8 of Chapter 2.
3. The vector
$$\begin{bmatrix} 4/7 \\ 3/7 \end{bmatrix} = \frac{3}{7} \begin{bmatrix} 4/3 \\ 1 \end{bmatrix}$$
is also an eigenvector of A (or B) belonging to $\lambda = 1$ since it too belongs to $\mathcal{E}_2(A)$.

EXAMPLE 5.1.10. How do we find eigenvalues of a triangular matrix? Illustrate the method with $A = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$.

SOLUTION. Eigenvalues are just the roots of the characteristic equation $\det(\lambda I - A) = 0$. Notice that $-A$ is triangular if A is. Also, the only entries in $\lambda I - A$ that are any different from the entries of $-A$ are the diagonal entries, which change from $-a_{ii}$ to $\lambda - a_{ii}$. Therefore, $\lambda I - A$ is triangular if A is. We already know that the determinant of a triangular matrix is easy to compute: just form the product of the diagonal entries. Therefore, the roots of the characteristic equation are the solutions to

$$0 = \det(\lambda I - A) = (\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{nn})$$

that is, $\lambda = a_{11}, a_{22}, \ldots, a_{nn}$. In other words, for a triangular matrix the eigenvalues are simply the diagonal elements! In particular, for the example A given above, we see with no calculations that the eigenvalues are $\lambda = 2, 1, -1$. $\square$

Notice, by the way, that we don't quite get off the hook in the preceding example if we are required to find the eigenvectors. It will still be some work to compute each of the relevant null spaces, but much less than it would take for a general matrix.

Example 5.1.10 can be used to illustrate another very important point. The reduced row echelon form of the matrix of that example is clearly the identity matrix I_3. This matrix has eigenvalues $1, 1, 1$, which are *not* the same as the eigenvalues of A (would that eigenvalue calculations were so easy!). In fact, a single elementary row operation on a matrix can change the eigenvalues. For example, simply multiply the first row of A above by $\frac{1}{2}$. This point warrants a warning, since it is the source of a fairly common mistake.

Caution: The eigenvalues of a matrix A and the matrix EA, where E is an elementary matrix, need not be the same.

EXAMPLE 5.1.11. Find an eigensystem for the matrix $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

SOLUTION. For eigenvalues, compute the roots of the equation

$$\begin{aligned} 0 = \det(A - \lambda I) &= \det \begin{bmatrix} 1-\lambda & -1 \\ 1 & 1-\lambda \end{bmatrix} \\ &= (1-\lambda)^2 - (-1) \\ &= \lambda^2 - 2\lambda + 2. \end{aligned}$$

Now we have a little problem. Do we allow complex numbers? If not, we are stuck because the roots of this equation are

$$\lambda = \frac{-(-2) \pm \sqrt{(-2)^2 - 4 \cdot 2}}{2} = 1 \pm i$$

In other words, if we did not enlarge our field of scalars to the complex numbers, we would have to conclude that there are *no* eigenvalues or eigenvectors! Somehow, this doesn't seem like a good idea. It is throwing information away. Perhaps it comes as no surprise that complex numbers would eventually figure into the eigenvalue story. After all, finding eigenvalues is all about solving polynomial equations, and complex numbers were invented to overcome the inability of real numbers to provide solutions to all polynomial equations. Let's allow complex numbers as the scalars. Now our eigenspace calculations are really going on in the complex space $\mathbb{C}^2$ instead of $\mathbb{R}^2$.

$\lambda = 1+i$: Then $A-(1+i)I = \begin{bmatrix} 1-(1+i) & -1 \\ 1 & 1-(1+i) \end{bmatrix} = \begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix}$ and row reduction gives (recall that $1/i = -i$)

$$\begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \xrightarrow[E_1(1/(-i))]{E_{21}(-i)} \begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix}$$

so the general solution is

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} iz_2 \\ z_2 \end{bmatrix} = z_2 \begin{bmatrix} i \\ 1 \end{bmatrix}$$

Therefore a basis of $\mathcal{E}_{1+i}(A)$ is $\{(i,1)\}$.

$\lambda = 1-i$: Then $A-(1-i)I = \begin{bmatrix} 1-(1-i) & -1 \\ 1 & 1-(1-i) \end{bmatrix} = \begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix}$ and row reduction gives (remember that $1/i = -i$)

$$\begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix} \xrightarrow[E_1(1/i)]{E_{21}(i)} \begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix}$$

so the general solution is

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} -iz_2 \\ z_2 \end{bmatrix} = z_2 \begin{bmatrix} -i \\ 1 \end{bmatrix}$$

Therefore a basis of $\mathcal{E}_{1+i}(A)$ is $\{(-i,1)\}$. □

In view of the previous example, we are going to adopt the following practice: unless otherwise stated, if the eigenvalue calculation leads us to complex numbers, we take the point of view that the field of scalars should be enlarged to include the complex numbers and the eigenvalues in question.

Multiplicity of Eigenvalues

The following example presents yet another curiosity about eigenvalues and vectors.

EXAMPLE 5.1.12. Find an eigensystem for the matrix $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$.

SOLUTION. Here the eigenvalues are easy. This matrix is triangular, so they are $\lambda = 2, 2$. Now for eigenvectors.

$\lambda = 2$: Then $A - 2I = \begin{bmatrix} 2-2 & 1 \\ 0 & 2-2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and row reduction is not necessary here. Notice that the variable x_1 is free here while x_2 is bound. The general solution is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Therefore a basis of $\mathcal{E}_2(A)$ is $\{(1,0)\}$. □

The manner in which we list the eigenvalues in this example is intentional. The number 2 occurs twice on the diagonal, suggesting that it should be counted twice. As a matter of fact, $\lambda = 2$ is a root of the characteristic equation $(\lambda - 2)^2 = 0$ of multiplicity 2. Yet there is a curious mismatch here. In all of our examples to this point, we have been able to come up with as many eigenvectors as eigenvalues, namely the size of the matrix if we allow complex numbers. In this case there is a deficiency in the number of eigenvectors, since there is only one eigenspace and it is one dimensional. Is this a failing entirely due to the occurrence of multiple eigenvalues? The answer is no. The situation is a bit more complicated, as the following example shows.

EXAMPLE 5.1.13. Discuss the eigenspace corresponding to the eigenvalue $\lambda = 2$ for these two matrices for these two matrices

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

SOLUTION. Notice that each of these matrices has eigenvalues $\lambda = 1, 2, 2$. Now for the eigenspace $\mathcal{E}_2(A)$.

(a) For this eigenspace calculation we have

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

and row reduction gives

$$\begin{bmatrix} 0 & 1 & 2 \\ 0 & -1 & -2 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[E_{21}(1)]{} \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

so that free variables are x_1, x_3 and the general solution is

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

Thus a basis for $\mathcal{E}_2(A)$ is $\{(1,0,0), (0,-2,1)\}$. Notice that in this case we get as many independent eigenvectors as the eigenvalue $\lambda = 2$ occurs.

(b) For this eigenspace calculation we have

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

and row reduction gives

$$\begin{bmatrix} 0 & 1 & 2 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[E_{21}(1)]{} \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[\substack{E_{12}(2) \\ E_2(-1)}]{} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

so that the only free variable is x_1 and the general solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Thus a basis for $\mathcal{E}_2(A)$ is $\{(1,0,0)\}$. Notice that in this case we don't get as many independent eigenvectors as the eigenvalue $\lambda = 2$ occurs. □

This example shows that there are two kinds of "multiplicities" of an eigenvector. On the one hand there is the number of times that the eigenvalue occurs as a root of the characteristic equation. On the other hand there is the dimension of the corresponding eigenspace. One of these is algebraic in nature, the other is geometric. Here are the appropriate definitions.

Algebraic and Geometric Multiplicity

DEFINITION 5.1.14. Let λ be a root of the characteristic equation $\det(\lambda I - A) = 0$. The *algebraic* multiplicity of λ is the multiplicity of λ as a root of the characteristic equation. The *geometric* multiplicity of λ is the dimension of the space $\mathcal{E}_\lambda(A) = \mathcal{N}(\lambda I - A)$.

We categorize eigenvalues as simple or repeated, according to the following definition.

DEFINITION 5.1.15. The eigenvalue λ of A is said to be *simple* if its algebraic multiplicity is 1, that is, the number of times it occurs as a root of the characteristic equation is 1. Otherwise, the eigenvalue is said to be *repeated*.

In Example 5.1.13 we saw that the repeated eigenvalue $\lambda = 2$ has algebraic multiplicity 2 in both (a) and (b), but geometric multiplicity 2 in (a) and 1 in (b). What can be said in general? The following theorem summarizes the facts. In particular, part 2 says that *algebraic multiplicity is always greater than or equal to geometric multiplicity*. Part 1 is immediate since a polynomial of degree n has n roots, counting complex roots and multiplicities. We defer the proof of part 2 to the next section.

THEOREM 5.1.16. *Let A be an $n \times n$ matrix with characteristic polynomial $p(\lambda) = \det(\lambda I - A)$. Then:*

1. *The number of eigenvalues of A, counting algebraic multiplicities and complex numbers, is n.*
2. *For each eigenvalue λ of A, if $m(\lambda)$ is the algebraic multiplicity of λ, then*
$$1 \leq \dim \mathcal{E}_\lambda(A) \leq m(\lambda)$$

Now when we wrote that each of the matrices of Example 5.1.13 has eigenvalues $\lambda = 1, 2, 2$, what we intended to indicate was a complete listing of the eigenvalues of the matrix, counting algebraic multiplicities. In particular, $\lambda = 1$ is a simple eigenvalue of the matrices, while $\lambda = 2$ is not. The geometric multiplicities of (a) are identical to the algebraic in (a) but not in (b). The latter kind of matrix is harder to deal with than the former. Following a time honored custom of mathematicians, we call the more difficult matrix by a less than flattering name, namely, "defective."

DEFINITION 5.1.17. A *defective* matrix is one for which the sum of the geometric multiplicities is strictly less than the sum of the algebraic multiplicities.

Notice that the sum of the algebraic multiplicities of an $n \times n$ matrix is the size n of the matrix. This is due to the fact that the characteristic polynomial of the matrix has degree n, hence exactly n roots, counting multiplicities.

5.1 Exercises

1. Find eigenvalues for these matrices:

(a) $\begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 6 & 2 \end{bmatrix}$ (c) $\begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{bmatrix}$ (d) $\begin{bmatrix} 2 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 2 \end{bmatrix}$

(e) $\begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$ (f) $\begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}$ (g) $\begin{bmatrix} 0 & 2 \\ 2 & 1 \end{bmatrix}$ (h) $\begin{bmatrix} 1+i & 3 \\ 0 & i \end{bmatrix}$

2. Find eigensystems for the matrices of Exercise 1 and specify the algebraic and geometric multiplicity of each eigenvalue.

3. You are given that the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has eigenvalues $1, -1$ and respective eigenvectors $(1, 1), (1, -1)$. Use Theorem 5.1.5 to determine an eigensystem for $\begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix}$ without further eigensystem calculations.

4. The *trace* of a matrix A is the sum of all the diagonal entries of the matrix and denoted $\operatorname{tr} A$. Find the trace of each matrix in Exercise 1 and verify that it is the sum of the eigenvalues of the matrix.

5. Let

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

be a general 2×2 matrix.

(a) Compute the characteristic polynomial of A and find its roots, i.e., the eigenvalues of A.

(b) Show that the sum of the eigenvalues is the trace of A.

(c) Show that the product of the eigenvalues is the determinant of A.

6. Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$ and show that A and A^T have different eigenvectors but the same eigenvalues.

7. Show that the matrix $A = \begin{bmatrix} 1 & i \\ i & 2 \end{bmatrix}$ is symmetric and that its eigenvalues are complex.

8. Show that for any square matrix A, the matrices A and A^T have the same eigenvalues.

9. Show that if $\mathbf{x}$ is an eigenvector for the matrix A belonging to the eigenvalue λ, then so is $c\mathbf{x}$ for any scalar $c \neq 0$.

10. Let A be a matrix whose eigenvalues are all less than 1 in absolute value. Show that every eigenvalue of $I - A$ is nonzero and deduce that $I - A$ is invertible.

11. Prove that if A is invertible and λ is an eigenvalue of A, then $1/\lambda$ is an eigenvalue of A^{-1}.

12. Let A be any square real matrix and show that the eigenvalues of A^TA are all nonnegative.

13. Show that if λ is an eigenvalue of an orthogonal matrix P, then $|\lambda| = 1$.

14. Let T_k be the $k \times k$ tridiagonal matrix whose diagonal entries are 2 and off-diagonal nonzero entries are -1. Use a MAS or CAS (MAS would probably be better) to build an array y of length 30 whose kth entry is the minimum of the absolute value of the eigenvalues of T_{k+1}. Plot this array. Use the graph as a guide and try to approximate $y(k)$ as a simple function of k.

15. Show that if B is a real symmetric positive definite matrix, then the eigenvalues of B are positive.

16. Let A be a real matrix and $(\lambda, \mathbf{x})$ an eigenpair for A.

(a) Show that $(\overline{\lambda}, \overline{\mathbf{x}})$ is also an eigenpair for A.

(b) Given that $A = \begin{bmatrix} 2 & -2 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 2 \end{bmatrix}$ and that $(2, (-1, 0, 1))$ and $(1+i, (1+i, 1, 0))$ are eigenpairs, with no further calculations exhibit an eigensystem for the matrix A and a matrix P for which $P^{-1}AP$ is diagonal.

(c) Deduce from part (a) that the real quadratic $\lambda^2 - 2\Re(\lambda) + |\lambda|^2$ is a factor of the characteristic polynomial of A.

17. Let A and B be matrices of the same size with eigenvalues α and β, respectively. Show by example that it is false to conclude that $\alpha + \beta$ is an eigenvalue of $A + B$ or that $\alpha\beta$ is an eigenvalue of AB.

18. Show that A and A^T have the same eigenvalues.

19. Show that if A and B are the same size, then AB and BA have the same eigenvalues. Hint: Deal with the 0 eigenvalue separately. If λ is an eigenvalue of AB,multiply the equation $AB\mathbf{x} = \lambda\mathbf{x}$ on the left by B.

5.2. Similarity and Diagonalization

Diagonalization and Matrix Powers

Eigenvalues: why are they? This is a good question and the justification for their existence and study could go on and on. We will try to indicate their importance by focusing on one special class of problems, namely, *discrete linear dynamical systems*. Here is the definition of such a system.

DEFINITION 5.2.1. A *discrete linear dynamical system* is a sequence of vectors $\mathbf{x}^{(k)}$, $k = 0, 1, \ldots$, called *states*, which is defined by an initial vector $\mathbf{x}^{(0)}$ and by the rule

$$\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)}, \qquad k = 0, 1, \ldots$$

where A is a given fixed square matrix, called the *transition matrix* of the system.

We have seen examples of this kind of system before, namely in Markov chains and difference equations. Evidently, the entire sequence of state vectors is determined by the matrix A and the initial state $\mathbf{x}^{(0)}$. Here is the sort of question that we would like to answer: when is it the case that there is a limiting vector $\mathbf{x}$ for this sequence of vectors and, if so, how does one compute this vector? The answers to these questions will explain the behavior of the Markov chain that was introduced in Example 2.3.4 of Chapter 2.

Discrete Dynamical System

If there is such a limiting vector $\mathbf{x}$ for a Markov chain, we saw in Example 3.4.8 of Chapter 3 how to proceed: find the null space of the matrix $I - A$, that is, the set of all solutions to the system $(I - A)\mathbf{x} = 0$. However, the question of whether or not all initial states $\mathbf{x}^{(0)}$ lead to this limiting vector is a more subtle issue which requires the insights of the next section. We've already done some work on this problem. We saw in Section 2.3 that the entire sequence of vectors is uniquely determined by the initial vector and the transition matrix A in the explicit formula

$$\mathbf{x}^{(k)} = A^k\mathbf{x}^{(0)}.$$

Before proceeding further, let's consider another example that will indicate why we would be interested in limiting vectors.

EXAMPLE 5.2.2. By some unfortunate accident a new species of frog has been introduced into an area where it has too few natural predators. In an attempt to restore the ecological balance, a team of scientists is considering introducing a species of bird which feeds on this frog. Experimental data suggests that the population of frogs and birds from one year to the next can be modeled by linear relationships. Specifically, it has been found that if the quantities F_k and B_k represent the populations of the frogs and birds in the kth year, then

$$\begin{aligned} B_{k+1} &= 0.6B_k + 0.4F_k \\ F_{k+1} &= -rB_k + 1.4F_k \end{aligned}$$

Here the positive number r is a kill rate which measures the consumption of frogs by birds. It varies with the environment, depending on factors such as the availability of other food for the birds, etc. Experimental data suggests in the environment where the birds are to be introduced, $r = 0.35$. The question is this: in the long run, will the introduction of the birds reduce or eliminate the frog population growth?

SOLUTION. The discrete dynamical system concept introduced in the preceding discussion fits this situation very nicely. Let the population vector in the kth year be $\mathbf{x}^{(k)} = (B_k, F_k)$. Then the linear relationship above becomes

$$\begin{bmatrix} B_{k+1} \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} 0.6 & 0.4 \\ -0.35 & 1.4 \end{bmatrix} \begin{bmatrix} B_k \\ F_k \end{bmatrix}$$

which is a discrete linear dynamical system. Notice that this is different from the Markov chains we studied earlier, since one of the entries of the coefficient matrix is negative. Before we can finish solving this example we need to have a better understanding of discrete dynamical systems and the relevance of eigenvalues. □

Let's try to understand how state vectors change in the general discrete dynamical system. We have $\mathbf{x}^{(k)} = A^k\mathbf{x}^{(0)}$. So, to understand how a dynamical system works, what we really need to know is how the powers of the transition matrix A behave. But in general, this is very hard!

Here is an easy case we can handle: what if $A = [a_{ij}]$ is diagonal? Since we'll make extensive use of diagonal matrices, let's use the following notation.

Notation: The matrix $\operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ is the $n \times n$ diagonal matrix with entries $\lambda_1, \lambda_2, \ldots, \lambda_n$ down the diagonal.

For example,

$$\operatorname{diag}\{\lambda_1, \lambda_2, \lambda_3\} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

By matching up the ith row and jth column of A we see that the only time we could have a nonzero entry in A^2 is when $i = j$, and in that case the entry is a_{ii}^2. A similar argument applies to any power of A. In summary, we have this handy

THEOREM 5.2.3. *If $D = \operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, then $D^k = \operatorname{diag}\{\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k\}$, for all positive integers k.*

Just as an aside, this theorem has a very interesting consequence. We have seen in some exercises that if $f(x) = a_0 + a_1x + \ldots + a_nx^n$ is a polynomial, we can evaluate $f(x)$ at the square matrix A as long as we understand that the constant term a_0 is evaluated as a_0I. This notion of $f(A)$ has some important applications. In the case of a diagonal A, the following fact reduces evaluation of $f(A)$ to a sequence of scalar calculations.

COROLLARY 5.2.4. *If $D = \operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$and $f(x)$ is a polynomial, then*

$$f(D) = \operatorname{diag}\{f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)\}$$

PROOF. Simply observe that $f(D) = a_0I + a_1D + \ldots + a_nD^n$, apply the preceding theorem to each monomial and add diagonal terms up. □

Now for the powers of a more general A. For ease of notation, let's consider a 3×3 matrix A. What if we could find three linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$? We would have

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \;\; A\mathbf{v}_2 = \lambda_2\mathbf{v}_2, \;\; A\mathbf{v}_3 = \lambda_3\mathbf{v}_3$$

or

$$A[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] \operatorname{diag}\{\lambda_1, \lambda_2, \lambda_3\}.$$

Now set

$$P = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$$

and

$$D = \operatorname{diag}\{\lambda_1, \lambda_2, \lambda_3\}$$

Then P is invertible since the columns of P are linearly independent. (Remember that any nonzero solution to $A\mathbf{x} = 0$ would give rise to a nontrivial linear combination of the column of A that sums to 0.) Moreover, the equation $AP = PD$, if multiplied on the left by P^{-1}, gives the equation

$$P^{-1}AP = D$$

This is a beautiful equation, because it makes the powers of A simple to understand. The procedure we just went through is reversible as well. In other words, if P is a given invertible matrix such that $P^{-1}AP = D$, then we can obtain that $AP = PD$, identify the columns of P by the equation $P = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$ and conclude that the columns of P are linearly independent eigenvectors of A. We make the following definition and follow it with a simple but key theorem relating similar matrices.

DEFINITION 5.2.5. A matrix A is said to be *similar* to a matrix B if there exists an invertible matrix P such that Similar Matrices

$$P^{-1}AP = B$$

A simple size check shows that similar matrices have to be square and of the same size. Furthermore, if A is similar to B, then B is similar to A. To see this, suppose that $P^{-1}AP = B$ and multiply by P on the left and P^{-1} on the right to obtain that

$$A = PP^{-1}APP^{-1} = PBP^{-1} = (P^{-1})^{-1}BP^{-1}$$

Similar matrices have much in common, as the following theorem shows.

THEOREM 5.2.6. *Suppose that A is similar to B, say $P^{-1}AP = B$. Then:*

1. *For every positive integer k,*
$$B^k = P^{-1}A^kP$$
2. *The matrices A and B have the same characteristic polynomial, hence the same eigenvalues.*

PROOF. We see that successive terms $P^{-1}P$ cancel out in the k-fold product

$$B^k = (P^{-1}AP)(P^{-1}AP)\cdots(P^{-1}AP)$$

to give that

$$B^k = P^{-1}A^kP$$

This proves (1). For (2), remember that the determinant distributes over products, so that we can pull this clever little trick:

$$\begin{aligned}
\det(\lambda I - B) &= \det(\lambda P^{-1}IP - P^{-1}AP) \\
&= \det(P^{-1}(\lambda I - A)P) \\
&= \det(P^{-1})\det(\lambda I - A)\det(P) \\
&= \det(\lambda I - A)\det(P^{-1}P) \\
&= \det(\lambda I - A).
\end{aligned}$$

This proves (2). □

Now we can see the significance of the equation $P^{-1}AP = D$, where D is diagonal. It follows from this equation that for any positive integer k, we have $P^{-1}A^kP = D^k$, so multiplying on the left by P and on the right by P^{-1} yields

$$A^k = PD^kP^{-1} \tag{5.2.1}$$

As we have seen, the term PD^kP^{-1} is easily computed.

EXAMPLE 5.2.7. Apply the results of the preceding discussion to the matrix in part (a) of Example 5.1.13.

SOLUTION. The eigenvalues of this problem are $\lambda = 1, 2, 2$. We already found the eigenspace for $\lambda = 2$. Denote the two basis vectors by $\mathbf{v}_1 = (1, 0, 0)$ and $\mathbf{v}_2 = (0, -2, 1)$. For $\lambda = 1$, apply Gauss-Jordan elimination to the matrix

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

and we can obviously reduce this matrix by using the pivot in the third row to

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

which gives a general eigenvector of the form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_2 \\ 0 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

Hence the eigenspace $E_1(A)$ has basis $\{(-1, 1, 0)\}$. Now set $\mathbf{v}_3 = (-1, 1, 0)$. Form the matrix

$$P = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = \begin{bmatrix} 1 & 0 & -1 \\ 0 & -2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

This matrix is nonsingular since $\det P = -1$, and a calculation which we leave to the reader shows that

$$P^{-1} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

The discussion of the first part of this section shows us that

$$P^{-1}AP = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = D$$

As we have seen, this means that for any positive integer k, we have

$$\begin{aligned} A^k &= PD^kP^{-1} \\ &= \begin{bmatrix} 1 & 0 & -1 \\ 0 & -2 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2^k & 0 & 0 \\ 0 & 2^k & 0 \\ 0 & 0 & 1^k \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 2^k & 2^k - 1 & 2^{k+1} - 2 \\ 0 & 1 & -2^{k+1} + 2 \\ 0 & 0 & 2^k \end{bmatrix} \end{aligned}$$

This is the formula we were looking for. It's *much* easier than calculating A^k directly! □

This example showcases a very nice calculation. Given a general matrix A, when can we pull off the same calculation? First, let's give the favorable case a name.

DEFINITION 5.2.8. The matrix A is *diagonalizable* if it is similar to a diagonal matrix, that is, there is an invertible matrix P and diagonal matrix D such that $P^{-1}AP = D$. In this case we say that P is a *diagonalizing matrix* for A or that P *diagonalizes* A.

The question is, can we be more specific about when a matrix is diagonalizable? We can. As a first step, notice that the calculations that we began the section with can easily be written in terms of an $n \times n$ matrix instead of 3×3. What these calculations prove is the following basic fact.

THEOREM 5.2.9. *The $n \times n$ matrix A is diagonalizable if and only if there exists a linearly independent set of eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of A, in which case $P = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$ is a diagonalizing matrix for A.* Diagonalization Theorem

Can we be more specific about when a linearly independent set of eigenvectors exists? Actually, we can. Clues about what is really going on can be gleaned from a re-examination of Example 5.1.13.

EXAMPLE 5.2.10. Apply the results of the preceding discussion to the matrix in part (b) of Example 5.1.13 or explain why they fail to apply.

SOLUTION. The eigenvalues of this problem are $\lambda = 1, 2, 2$. We already found the eigenspace for $\lambda = 2$. Denote the single basis vector of $\mathcal{E}_2(A)$ by $\mathbf{v}_1 = (1, 0, 0)$. For $\lambda = 1$, apply Gauss-Jordan elimination to the matrix

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

and we can obviously reduce this matrix by using the pivot in the second row to

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

which gives a general eigenvector of the form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_2 \\ 0 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}.$$

Hence the eigenspace $E_1(A)$ has basis $\{(-1,1,0)\}$. All we could come up with here is two eigenvectors. As a matter of fact, they are linearly independent since one is not a multiple of the other. But they aren't enough and there is no way to find a third eigenvector, since we have found them all! Therefore we have no hope of diagonalizing this matrix according to the diagonalization theorem. The real problem here is that A is defective, since the algebraic multiplicity of $\lambda = 2$ exceeds the geometric multiplicity of this eigenvalue. □

It would be very handy to have some working criterion for when we can manufacture linearly independent sets of eigenvectors. The next theorem gives us such a criterion.

THEOREM 5.2.11. *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be a set of eigenvectors of the matrix A such that corresponding eigenvalues are all distinct. Then the set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is linearly independent.*

PROOF. Suppose the set is linearly dependent. Discard redundant vectors until we have a smallest linearly dependent subset, say $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ is such a set with $\mathbf{v}_i$ belonging to λ_i. All the vectors have nonzero coefficients in a linear combination that sums to zero, for we could discard the ones that have zero coefficient in the linear combination and still have a linearly dependent set. So there is some linear combination of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_m\mathbf{v}_m = 0 \tag{5.2.2}$$

with each $c_j \neq 0$ and $\mathbf{v}_j$ belonging to eigenvalue λ_j. Multiply (5.2.2) by λ_1 to obtain the equation

$$c_1\lambda_1\mathbf{v}_1 + c_2\lambda_1\mathbf{v}_2 + \ldots + c_m\lambda_1\mathbf{v}_m = 0. \tag{5.2.3}$$

Next multiply (5.2.2) on the left by A to obtain

$$0 = A(c_1\lambda_1\mathbf{v}_1 + c_2\lambda_1\mathbf{v}_2 + \ldots + c_m\lambda_1\mathbf{v}_m) = c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 + \ldots + c_kA\mathbf{v}_k,$$

that is,

$$c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \ldots + c_k\lambda_m\mathbf{v}_m = 0. \tag{5.2.4}$$

Now subtract (5.2.4) from (5.2.3) to obtain

$$0\mathbf{v}_1 + c_2(\lambda_1 - \lambda_2)\mathbf{v}_2 + \ldots + c_k(\lambda_1 - \lambda_m)\mathbf{v}_m = 0$$

This is a new nontrivial linear combination (since $c_2(\lambda_1 - \lambda_2) \neq 0$) of fewer terms which contradicts our choice of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$. It follows that the original set of vectors must be linearly independent. □

Actually, a little bit more is true: if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is such that for any eigenvalue λ of A, the subset of all these vectors belonging to λ is linearly independent, then the conclusion of the theorem is valid. We leave this as an exercise. Here's an application of the theorem which is useful for many problems.

COROLLARY 5.2.12. *If the $n \times n$ matrix A has n distinct eigenvalues, then A is diagonalizable.*

PROOF. We can always find one nonzero eigenvector $\mathbf{v}_i$ for each eigenvalue λ_i of A. By the preceding theorem, the set $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is linearly independent. Thus A is diagonalizable by the Diagonalization theorem. □

Caution: Just because the $n \times n$ matrix A has fewer than n distinct eigenvalues, you may not conclude that it is not diagonalizable.

An example that illustrates this caution is part (a) of Example 5.1.13.

5.2 Exercises

1. Given each matrix A below, find a matrix P such that $P^{-1}AP$ is diagonal. Use this to deduce a formula for A^k.

$$\text{(a)} \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{(c)} \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$$

2. Determine if the following matrices are diagonalizable with a minimum of calculation.

$$\text{(a)} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} 1 & 3 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \qquad \text{(c)} \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$$

3. For each of the following matrices A find the characteristic polynomial $p(x)$ and evaluate $p(A)$. (This means that the matrix A replaces every occurrence of x and the constant term c_0 is replaced by $c_0 I$.)

$$\text{(a)} \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} 1 & 3 & 0 \\ -2 & 2 & 1 \\ 4 & 0 & -2 \end{bmatrix} \qquad \text{(c)} \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

4. Suppose that A is an invertible matrix which is diagonalized by the matrix P, that is, $P^{-1}AP = D$ is a diagonal matrix. Use this information to find a diagonalization for A^{-1}.

5. Adapt the proof of Theorem 5.2.11 to prove that if eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are such that for any eigenvalue λ of A, the subset of all these vectors belonging to λ is linearly independent, then the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly independent.

6. Suppose that the kill rate r of Example 5.2.2 is viewed as a variable positive parameter. There is a value of the number r for which the eigenvalues of the corresponding matrix are equal.

(a) Find this value of r and the corresponding eigenvalues by examining the characteristic polynomial of the matrix.

(b) Use the available MAS (or CAS) to determine experimentally the long term behavior of populations for the value of r found in (a). Your choices of initial states should include $[100, 1000]$.

7. The thirteenth century mathematician Leonardo Fibonacci discovered the sequence of integers $1, 1, 2, 3, 5, 8, \ldots$ called the *Fibonacci sequence*. These numbers have a way of turning up in many applications. They can be specified by the formulas

$$\begin{aligned} f_0 &= 1 \\ f_1 &= 1 \\ f_{n+2} &= f_{n+1} + f_n, \qquad n = 0, 1, \ldots. \end{aligned}$$

(a) Let $\mathbf{x}_n = (f_{n+1}, f_n)$ and show that these equations are equivalent to the matrix equations $\mathbf{x}_0 = (1,1)$ and $\mathbf{x}_{n+1} = A\mathbf{x}_n$, $n = 0, 1, \ldots$, where $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$

(b) Use part (a) and the diagonalization theorem to find an explicit formula for the nth Fibonacci number.

8. Calculate second, third and fourth powers of the matrix $J_\lambda(3) = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$. Based on these calculations, make a conjecture about the form of $J_\lambda(3)^k$, where k is any positive integer.

9. Show that any upper triangular matrix with constant diagonal is diagonalizable if and only if it is already diagonal. *Hint:* What diagonal matrix would such a matrix be similar to?

10. Let A be a 2×2 transition matrix of a Markov chain where A is not the identity matrix.

(a) Show that A can be written in the form $A = \begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix}$ for suitable real numbers $0 \le a, b \le 1$.

(b) Show that (b, a) and $(1, -1)$ are eigenvectors for A.

(c) Use (b) to diagonalize the matrix A and obtain a formula for the powers of A.

11. Show that if A is diagonalizable, then so is A^H.

12. Let A, B, P square matrices of the same size with P invertible.

(a) Show that $P^{-1}(A+B)P = P^{-1}AP + P^{-1}BP$

(b) Show that $P^{-1}ABP = (P^{-1}AP)(P^{-1}BP)$

(c) Use (a) and (b) to show that if $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ is a polynomial function, then $P^{-1}f(A)P = f(P^{-1}AP)$.

13. Prove the Cayley-Hamilton theorem for diagonalizable matrices; that is, show that if $p(x)$ is the characteristic polynomial of the diagonalizable matrix A, then A satisfies its characteristic equation, that is, $p(A) = 0$. *Hint:* You may find Exercise 12 and Corollary 5.2.4 very helpful.

14. Let A and B be matrices of the same size and suppose that A has no repeated eigenvalues. Show that $AB = BA$ if and only if A and B are simultaneously diagonalizable, that is, a single matrix P diagonalizes A and B. *Hint:* The diagonalization theorem and Exercise 24 are helpful.

5.3. Applications to Discrete Dynamical Systems

Now we have enough machinery to come to a fairly complete understanding of the discrete dynamical system

$$\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)}.$$

Diagonalizable Transition Matrix

Let us first examine the case that A is diagonalizable. So we assume that the $n \times n$ matrix A is diagonalizable and that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a complete linearly independent set of eigenvectors of A belonging to the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A. Let us further suppose that these eigenvalues are ordered so that

$$|\lambda_1| \le |\lambda_2| \le \ldots \le |\lambda_n|.$$

The eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ form a basis of $\mathbb{R}^n$ or $\mathbb{C}^n$, whichever is appropriate. In particular, we may write $\mathbf{x}^{(0)}$ as a linear combination of these vectors, say

$$\mathbf{x}^{(0)} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n. \tag{5.3.1}$$

Now we can see what the effect of multiplication by A is:

$$\begin{aligned} A\mathbf{x}^{(0)} &= A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) \\ &= c_1(A\mathbf{v}_1) + c_2(A\mathbf{v}_2) + \cdots + c_n(A\mathbf{v}_n) \\ &= c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_n\lambda_n\mathbf{v}_n. \end{aligned}$$

Now apply A on the left repeatedly and we see that

$$\mathbf{x}^{(k)} = A^k\mathbf{x}^{(0)} = c_1\lambda_1^k\mathbf{v}_1 + c_2\lambda_2^k\mathbf{v}_2 + \cdots + c_n\lambda_n^k\mathbf{v}_n. \tag{5.3.2}$$

□

Equation 5.3.2 is the key to understanding how the state vector changes in a discrete dynamical system. Now we can see clearly that it is the size of the eigenvalues that governs the growth of successive states. Because of this fact, a handy quantity that can be associated with a matrix A (whether it is diagonalizable or not) is the so-called *spectral radius* of A, which we denote by $\rho(A)$. This number is defined by the formula

$$\rho(A) = \max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_n|\}.$$

That is, $\rho(A)$ is the largest absolute value of the eigenvalues of A. We summarize a few of the conclusions that can be drawn in terms of the spectral radius and dominant eigenvalues.

THEOREM 5.3.1. *Let the transition matrix for a discrete dynamical system be the $n \times n$ diagonalizable matrix A as described above. Let $\mathbf{x}^{(0)}$ be an initial state vector given as in Equation 5.3.1. Then the following are true:*

1. *If $\rho(A) < 1$, then $\lim_{k\to\infty} \mathbf{x}^{(k)} = 0$.*
2. *If $\rho(A) = 1$, then the sequence of norms $\{\left\|\mathbf{x}^{(k)}\right\|\}$ is bounded.*

3. *If* $\rho(A) = 1$ *and the only eigenvalues* λ *of* A *with* $|\lambda| = 1$ *are* $\lambda = 1$, *then* $\lim_{k\to\infty} \mathbf{x}^{(k)}$ *is an element of* $\mathcal{E}_1(A)$, *hence either an eigenvector or* 0.
4. *If* $\rho(A) > 1$, *then for some choices of* $\mathbf{x}^{(0)}$ *we have* $\lim_{k\to\infty} \left\|\mathbf{x}^{(k)}\right\| = \infty$.

PROOF. Suppose that $\rho(A) < 1$. Then for all i, $\lambda_i^k \to 0$ as $k \to \infty$, so we see from Equation 5.3.2 that $\mathbf{x}^{(k)} \to 0$ as $k \to \infty$, which is what (1) says. Next suppose that $\rho(A) = 1$. Then take norms of Equation 5.3.2 to obtain that, since each $|\lambda_i| \leq 1$,

$$\begin{aligned}\left\|\mathbf{x}^{(k)}\right\| = \left\|A^k\mathbf{x}^{(0)}\right\| &= \left\|c_1\lambda_1^k\mathbf{v}_1 + c_2\lambda_2^k\mathbf{v}_2 + \cdots + c_n\lambda_n^k\mathbf{v}_n\right\| \\ &\leq |\lambda_1|^k \|c_1\mathbf{v}_1\| + |\lambda_2|^k \|c_2\mathbf{v}_2\| + \cdots + |\lambda_n|^k \|c_n\mathbf{v}_n\| \\ &\leq \|c_1\mathbf{v}_1\| + \|c_2\mathbf{v}_2\| + \cdots + \|c_n\mathbf{v}_n\|.\end{aligned}$$

Therefore the sequence of norms $\left\|\mathbf{x}^{(k)}\right\|$ is bounded by a constant that only depends on $\left\|\mathbf{x}^{(0)}\right\|$, which proves (2). The proof of (3) follows from inspection of (5.3.2): Observe that the eigenvalue powers $\lambda_j^k = 1$ if $\lambda = 1$ and otherwise the powers tend to zero since all other eigenvalues are less than 1 in absolute value. Hence if any coefficient c_j of an eigenvector $\mathbf{v}_j$ corresponding to 1 is not zero, the limiting vector is an eigenvector corresponding to $\lambda = 1$. Otherwise, the coefficients all tend to 0 and the limiting vector is 0. Finally, if $\rho(A) > 1$, then for $\mathbf{x}^{(0)} = c_n\mathbf{v}_n$, we have that $\mathbf{x}^{(k)} = c_n\lambda_n^k\mathbf{v}_n$. However, $|\lambda_n| > 1$, so that $\left|\lambda_n^k\right| \to \infty$, as $k \to \infty$, from which the desired conclusion for (4) follows. □

We should note that the cases of the preceding theorem are not quite exhaustive. One possibility that is not covered is the case that $\rho(A) = 1$ and A has other eigenvalues of absolute value 1. In this case the sequence of vectors $\mathbf{x}^{(k)}$, $k = 0, 1, \ldots$, is bounded in norm but need not converge to anything. An example of this phenomenon is given in Example 5.3.4

EXAMPLE 5.3.2. Apply the preceding theory to the population of Example 5.2.2.

SOLUTION. We saw in this example that the transition matrix is

$$A = \begin{bmatrix} 0.6 & 0.4 \\ -0.35 & 1.4 \end{bmatrix}.$$

The characteristic equation of this matrix is

$$\begin{aligned}\det \begin{bmatrix} 0.6-\lambda & 0.4 \\ -0.35 & 1.4-\lambda \end{bmatrix} &= (0.6-\lambda)(1.4-\lambda) + 0.35 \cdot 0.4 \\ &= \lambda^2 - 2\lambda + 0.84 + 0.14 \\ &= \lambda^2 - 2\lambda + 0.98\end{aligned}$$

whence we see that the eigenvalues of A are

$$\begin{aligned}\lambda &= 1.0 \pm \sqrt{4 - 3.92}/2 \\ &\approx 0.85858,\ 1.1414\end{aligned}$$

A calculation which we leave to the reader also shows that the eigenvectors of A corresponding to these eigenvalues are approximately $\mathbf{v}_1 = (.8398, .54289)$ and $\mathbf{v}_2 = (1.684, 2.2794)$, respectively. Since $\rho(A) \approx 1.1414 > 1$, it follows from (1) of Theorem 5.3.1 that for every initial state except a multiple of $\mathbf{v}_1$, the limiting state will grow without bound. Now if we imagine an initial state to be a random choice of values for the coefficients c_1 and c_2, we see that the probability of selecting $c_2 = 0$, is for all practical purposes, 0. Therefore, with probability 1, we will make a selection with $c_2 \neq 0$, from which it follows that the subsequent states will tend to arbitrarily large multiples of the vector $\mathbf{v}_2 = (1.684, 2.2794)$.

Finally, we can offer some advice to the scientists who are thinking of introducing a predator bird to control the frog population of this example: don't do it! Almost any initial distribution of birds and frogs will result in a population of birds and frogs that grows without bound. Therefore, we will be stuck with both non-indigenous frogs *and* birds. To drive the point home, start with a population of $10,000$ frogs and 100 birds. In 20 years we will have a population state of

$$\begin{bmatrix} 0.6 & 0.4 \\ -0.35 & 1.4 \end{bmatrix}^{20} \begin{bmatrix} 100 \\ 10,000 \end{bmatrix} \approx \begin{bmatrix} 197,320 \\ 267,550 \end{bmatrix}$$

In view of our eigensystem analysis, we know that these numbers are no fluke. Almost any initial population will grow similarly. The conclusion is that we should try another strategy or perhaps leave well enough alone in this ecology. □

EXAMPLE 5.3.3. Apply the preceding theory to the Markov chain Example 2.3.4 of Chapter 2.

SOLUTION. Recall that this example led to a Markov chain whose transition matrix is given by

$$A = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}.$$

Conveniently, we have already computed the eigenvalues and vectors of $10A$ in Example 5.1.9. There we found eigenvalues $\lambda = 3, 10$, with corresponding eigenvectors $\mathbf{v}_1 = (1, -1)$ and $\mathbf{v}_2 = (4/3, 1)$, respectively. It follows from Example 5.1.9 that the eigenvalues of A are $\lambda = 0.3,\ 1$, with the same eigenvectors. Therefore 1 is the dominant eigenvalue. Any initial state will necessarily involve $\mathbf{v}_2$ nontrivially, since multiples of $\mathbf{v}_1$ are not probability distribution vectors (the entries are of opposite signs). Thus we may apply part 3 of Theorem 5.3.1 to conclude that for any initial state, the only possible nonzero limiting state vector is some multiple of $\mathbf{v}_2$. Which multiple? Since the sum of the entries of each state vector sum to 1, the same must be true of the initial vector. Since

$$\begin{aligned} \mathbf{x}^{(0)} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 &= c_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + c_2 \begin{bmatrix} 4/3 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} c_1 1 + c_2\,(4/3) \\ c_1(-1) + c_2 1 \end{bmatrix} \end{aligned}$$

we see that

$$\begin{aligned} 1 &= c_1 1 + c_2\,(4/3) + c_1(-1) + c_2 1 \\ &= c_2(7/3) \end{aligned}$$

so that $c_2 = 3/7$. Now use the facts that $\lambda_1 = 0.3$, $\lambda_2 = 1$ and Equation 5.3.2 with $n = 2$ to see that the limiting state vector is

$$\lim_{k\to\infty} c_1(0.3)^k\mathbf{v}_1 + c_2 1^k\mathbf{v}_2 = c_2\mathbf{v}_2 = \begin{bmatrix} 4/7 \\ 3/7 \end{bmatrix} \approx \begin{bmatrix} .57143 \\ .42857 \end{bmatrix}.$$

Compare this vector with the result obtained by direct calculation in Example 2.2.5. □

When do complex eigenvalues occur and what do they mean? In general, all we can say is that the characteristic polynomial of a matrix, even if it is real, may have complex roots. This is an unavoidable fact, but it can be instructive. To see how this is so, consider the following example.

EXAMPLE 5.3.4. Suppose that a discrete dynamical system has transition matrix $A = \begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix}$, where a is a positive real number. What can be said about the states $\mathbf{x}^{(k)}, k = 1, 2, \ldots$ if the initial state $\mathbf{x}^{(0)}$ is an arbitrary nonzero vector?

SOLUTION. The eigenvalues of A are $\pm ai$. Now if $a < 1$ then according to part 1 of Theorem 5.3.1 the limiting state is 0. Part 3 of that theorem cannot occur for our matrix A since 1 cannot be an eigenvalue. So suppose $a \geq 1$. Since the eigenvalues of A are distinct, there is an invertible matrix P such that

$$P^{-1}AP = D = \begin{bmatrix} ai & 0 \\ 0 & -ai \end{bmatrix}.$$

So we see from Equation 5.2.1 that

$$A^k = PD^kP^{-1} = P \begin{bmatrix} (ai)^k & 0 \\ 0 & (-ai)^k \end{bmatrix} P^{-1}$$

The columns of P are eigenvectors of A, hence complex. We may take real parts of the matrix D^k to get a better idea of what the powers of A do. Now $i = e^{i\frac{\pi}{2}}$, so we may use DeMoivre's formula to get

$$\Re((ai)^k) = a^k \cos(k\frac{\pi}{2}) = (-1)^{k/2}a^k \quad \text{if } k \text{ is even}$$

We know that $\mathbf{x}^k = A^k\mathbf{x}^0$. In view of the above equation, we see that the states $\mathbf{x}^k$ will oscillate around the origin. In the case that $a = 1$ we expect the states to remain bounded, but if $a > 1$, we expect the values to become unbounded and oscillate in sign. This oscillation is fairly typical of what happens when complex eigenvalues are present, though it need not be as rapid as in this example. □

Non-Diagonalizable Transition Matrix

How can a matrix be non-diagonalizable? All the examples we have considered so far suggest that non-diagonalizability is the same as being defective. Put another way, diagonalizable equals non-defective. This is exactly right, as the following shows.

THEOREM 5.3.5. *The matrix A is diagonalizable if and only if it is non-defective.*

PROOF. Suppose that the $n \times n$ matrix A is diagonalizable. According to the diagonalization theorem, there exists a complete linearly independent set of eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ of the matrix A. The number of these vectors belonging to a given eigenvalue λ of A is a number $d(\lambda)$ at most the geometric multiplicity of λ, since they form a basis of the eigenspace $\mathcal{E}_\lambda(A)$. Hence their number is at most the algebraic multiplicity $m(\lambda)$ of λ by Theorem 5.1.16. Since sum of all the numbers $d(\lambda)$ is n, as is the sum of all the algebraic multiplicities $m(\lambda)$, it follows that the sum of the geometric multiplicities must also be n. The only way for this to happen is that for each eigenvalue λ, we have that geometric multiplicity equals algebraic multiplicity. Thus, A is non-defective.

Conversely, if A is non-defective, we can produce $m(\lambda)$ linearly independent eigenvectors belonging to each eigenvalue λ. Assemble all of these vectors and we have n eigenvectors such that for any eigenvalue λ of A, the subset of all these vectors belonging to λ is linearly independent. Therefore, the entire set of eigenvectors is linearly independent by the remark following Theorem 5.2.11. Now apply the diagonalization theorem to obtain that A is diagonalizable. □

The last item of business in our examination of diagonalization is to prove part 2 of Theorem 5.1.16, which asserts:

For each eigenvalue μ of A, if $m(\mu)$ is the algebraic multiplicity of μ, then

$$1 \leq \dim \mathcal{E}_\mu(A) \leq m(\mu).$$

To see why this is true, suppose the eigenvalue μ has geometric multiplicity k and that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is a basis for the eigenspace $\mathcal{E}_\mu(A)$. We know from the Steinitz substitution theorem that this set can be expanded to a basis of the vector space $\mathbb{R}^n$ (or $\mathbb{C}^n$), say

$$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$$

Form the nonsingular matrix

$$S = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$$

Let

$$B = [S^{-1}A\mathbf{v}_{k+1}, S^{-1}A\mathbf{v}_{k+2}, \ldots, S^{-1}A\mathbf{v}_n] = \begin{bmatrix} F \\ G \end{bmatrix}$$

where F consists of the first k rows of B and G the remaining rows. Thus we obtain that

$$\begin{aligned} AS &= [A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_n] \\ &= [\mu\mathbf{v}_1, \mu\mathbf{v}_2, \ldots, \mu\mathbf{v}_k, A\mathbf{v}_{k+1}, \ldots, A\mathbf{v}_n] \\ &= S \begin{bmatrix} \mu I_k & F \\ 0 & G \end{bmatrix} \end{aligned}$$

Now multiply both sides on the left by S^{-1} and we have

$$C = S^{-1}AS = \begin{bmatrix} \mu I_k & F \\ 0 & G \end{bmatrix}$$

We see that the block upper triangular matrix C is similar to A. By part 2 of Theorem 5.2.6 we see that A and C have the same characteristic polynomial. However, the characteristic polynomial of C is

$$\begin{aligned} p(\lambda) &= \det\left(\lambda I_n - \begin{bmatrix} \mu I_k & F \\ 0 & G \end{bmatrix}\right) \\ &= \det\left(\begin{bmatrix} (\lambda-\mu)I_k & F \\ 0 & G - \lambda I_{n-k} \end{bmatrix}\right) \\ &= \det(\lambda-\mu)I_k \cdot \det\left(G - \lambda I_{n-k}\right) \\ &= (\lambda-\mu)^k \det\left(G - \lambda I_{n-k}\right) \end{aligned}$$

The product term above results from Exercise 11 of Section 2.6 of Chapter 2. It follows that the algebraic multiplicity of μ as a root of $p(\lambda)$ is at least as large as k, which is what we wanted to prove.

Our newfound insight into non-diagonalizable matrices is somewhat of a negative nature – they are defective. Unfortunately, this isn't much help in determining the behavior of discrete dynamical systems with a non-diagonalizable transition matrix. If matrices are not diagonalizable, what simple kind of matrix *are* they reducible to? There is a very nice answer to this question; this answer requires the notion of a Jordan block, which can be defined as a $d \times d$ matrix of the form

$$J_d(\lambda) = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$$

where the entries off the main diagonal and first super-diagonal are understood to be zeros. This matrix is very close to being a diagonal matrix. Its true value comes from the following classical theorem, the proof of which is somewhat beyond the scope of this text. We refer the reader to the textbooks [**7**] or [**6**] of the bibliography for a proof. These texts are an excellent references for higher level linear algebra and matrix theory.

Jordan Canonical Form Theorem

THEOREM 5.3.6. *Every matrix A is similar to a block diagonal matrix which consists of Jordan blocks down the diagonal. Moreover, these blocks are uniquely determined by A up to order.*

In particular, if $J = S^{-1}AS$, where J consists of Jordan blocks down the diagonal, we call J "the" Jordan canonical form of the matrix A. This is a slight abuse of language, since the order of occurrence of the Jordan blocks of J could vary. To fix ideas, let's consider an example.

EXAMPLE 5.3.7. Find all possible Jordan canonical forms for a 3×3 matrix A whose eigenvalues are $-2, 3, 3$.

SOLUTION. Notice that each Jordan block $J_d(\lambda)$ contributes d eigenvalues λ to the matrix. Therefore, there can be only one 1×1 Jordan block for the eigenvalue -2 and either two 1×1 Jordan blocks for the eigenvalue 3 or one 2×2 block for the eigenvalue 3. Thus, the possible Jordan canonical forms for A (up to order of blocks) are

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

□

Notice that if all Jordan blocks are 1×1, then the Jordan canonical form of a matrix is simply a diagonal matrix. Thus, another way to say that a matrix is diagonalizable is to say that its Jordan blocks are 1×1. In reference to the previous example, we see that if the matrix has the first Jordan canonical form, then it is diagonalizable, while if it has the second, it is non-diagonalizable.

Now suppose that the matrix A is a transition matrix for a discrete dynamical system and A is not diagonalizable. What can one say? For one thing, the Jordan canonical form can be used to recover part 1 of Theorem 5.3.1. Part 4 remains valid as well; the proof we gave does not depend on A being diagonalizable. Unfortunately, things are a bit more complicated as regards parts (2) and (3). In fact, they fail to be true, as the following example shows.

EXAMPLE 5.3.8. Let $A = J_2(1)$. Show how parts (2) and (3) of Theorem 5.3.1 fail to be true for this matrix.

SOLUTION. We check that

$$A^2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix},$$

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

and in general

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

Now take $\mathbf{x}^{(0)} = (0, 1)$ and we see that

$$\mathbf{x}^{(k)} = A^k \mathbf{x}^{(0)} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} k \\ 1 \end{bmatrix}$$

It follows that the norms $\left|\left|\mathbf{x}^{(k)}\right|\right| = \sqrt{k^2+1}$ are not a bounded sequence, so that part 2 of the theorem fails to be true. Also, the sequence of vectors $\mathbf{x}^{(k)}$ does not converge to any vector in spite of the fact that 1 is the largest eigenvalue of A. Thus (3) fails as well. □

In spite of this example, the news is not all negative. It can be shown by way of the Jordan canonical form that a weaker version of (3) holds: if $\rho(A) = 1$, 1 is the only eigenvalue of A of absolute value 1 *and* the eigenvalue 1 has multiplicity 1, then the conclusion of (3) in Theorem 5.3.1 holds.

5.3 Exercises

1. Which of the following matrices are diagonalizable? Do not carry out the diagonalization, but give reasons for your answers.

$$\text{(a)} \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}, \text{(b)} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}, \text{(c)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

$$\text{(d)} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{(e)} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \text{(f)} \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

2. For each matrix A below find an eigensystem of A and use this to produce an invertible matrix P and diagonal matrix D such that $P^{-1}AP = D$, where

$$\text{(a) } A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \text{(b) } A = \frac{1}{2}\begin{bmatrix} 3 & 2 \\ -4 & -3 \end{bmatrix}, \text{(c) } A = \frac{1}{2}\begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$$

3. Do the powers A^k tend to 0 as k tends to infinity for any of the matrices of Exercise 2?

4. You are given that a 5×5 matrix has eigenvalues $2, 2, 3, 3, 3$. What are the possible Jordan canonical forms for this matrix?

5. Compute by hand or calculator various power of the matrix $J_3(1) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. Make a conjecture about a general formula for the kth power of this matrix, based on these calculations.

In the next two exercises, we use the idea of a dominant eigenvalue, that is, an eigenvalue λ of the matrix A such that λ is a simple eigenvalue and $|\lambda| > |\mu|$ for every other eigenvalue μ of A.

6. If the eigenvalues of a 3 × 3 matrix A are one of the following lists, determine which is the dominant eigenvalue of A, if any.

(a) $\lambda = 1, 2, 2$ (b) $\lambda = 2, 2, -4$ (c) $\lambda = 5/4, 1 + i, 1 - i$

7. Suppose the transition matrix A of a discrete dynamical system has a dominant eigenvalue of 1. What conclusion can you extract from Theorem 5.3.1? Illustrate this conclusion with an example.

8. Part (3) of Theorem 5.3.1 suggests that two possible limiting values are possible. Use your CAS or MAS to carry out this experiment: Compute a random 2×1 vector and normalize it by dividing by its length. Let the resulting initial vector be $\mathbf{x}^{(0)} = (x_1, x_2)$ and compute the state vector $\mathbf{x}^{(20)}$ using the transition matrix A of Example 5.3.3. Do this for a large number of times (say 500) and keep count of the number of times $\mathbf{x}^{(20)}$ is close to $\mathbf{0}$, say $\|\mathbf{x}^{(20)}\| < 0.1$. Conclusions?

9. Use a CAS or MAS to construct a 3 × 10 table whose jth column is $A^j\mathbf{x}$, where $\mathbf{x} = (1, 1, 1)$ and $A = \begin{bmatrix} 10 & 17 & 8 \\ -8 & -13 & -6 \\ 4 & 7 & 4 \end{bmatrix}$. What can you deduce about the eigenvalues of A based on inspection of this table? Give reasons. Check your claims by finding the eigenvalues of A.

10. A species of bird can be divided into three age groups, say birds of age less than 2 years for group 1, birds of age between 2 and 4 years for group 2, and birds of age between 4 and 6 years for the third group. Assume that for all practical purposes, these birds have at most a 6 year life span. It is determined that the survival rates for group 1 and 2 birds are 50% and 25%, respectively. On the other hand, group 1 birds produce no

offspring, while group 2 birds average 4 offspring in any biennium (period of 2 years), and birds in group 3 average 3 offspring in a biennium. Model this bird population as a discrete dynamical system, where a state vector (p_1, p_2, p_3) for a given biennium means that there are p_1 birds in group 1, p_2 birds in group 2 and p_3 birds in group 3.

11. Compute the largest eigenvalue of the transition matrix for the model of Exercise 10. What does this suggest about the biennial growth rate of the bird population?

12. Let $A = J_3(\lambda)$, the Jordan block. Show that the *Cayley-Hamilton theorem* is valid for A, that is, $p(A) = 0$, where $p(x)$ is the characteristic polynomial of A.

13. The financial model of Example 2.3.12 gave rise to difference equation which was converted to the dynamical system $x^{(k+1)} = Ax^{(k)}$, where the transition matrix is given by

$$A = \begin{bmatrix} 1 & 0.06 & 0.12 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Use a CAS or MAS to calculate the eigenvalues of this matrix. Deduce that A is diagonalizable and determine the approximate growth rate from one state to the next, given a random initial vector. Compare the growth rate with a flat interest rate.

5.4. Orthogonal Diagonalization

Orthogonal and unitary matrices are particularly attractive when we have to deal with inverses. Recall that one situation which calls for an inverse matrix is that of diagonalization. For A is diagonalizable when there exists an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix. We are going to explore some very remarkable facts about Hermitian and real symmetric matrices. These matrices are diagonalizable, and moreover, diagonalization can be accomplished by a unitary (orthogonal if A is real) matrix. This means that $P^{-1}AP = P^H AP$ is diagonal. In this situation we say that the matrix A is *unitarily (orthogonally) diagonalizable*.

Eigenvalue of Hermitian Matrices

As a first step, we need to observe a curious property of Hermitian matrices. It turns out that their eigenvalues are guaranteed to be real, even if the matrix itself is complex. This is one reason that these matrices are so nice to work with.

THEOREM 5.4.1. *If A is a Hermitian matrix, then the eigenvalues of A are real.*

PROOF. Let λ be an eigenvalue of A with corresponding nonzero eigenvector $\mathbf{x}$, so that $A\mathbf{x} = \lambda\mathbf{x}$. Form the scalar $c = \mathbf{x}^H A\mathbf{x}$. We have that

$$\overline{c} = c^H = \left(\mathbf{x}^H A\mathbf{x}\right)^H = \mathbf{x}^H A^H (\mathbf{x}^H)^H = \mathbf{x}^H A\mathbf{x} = c$$

It follows that c is a real number. However, we also have that

$$c = \mathbf{x}^H \lambda\mathbf{x} = \lambda\mathbf{x}^H\mathbf{x} = \lambda\, ||\mathbf{x}||^2$$

so that $\lambda = c/\,||\mathbf{x}||^2$ is also real. □

EXAMPLE 5.4.2. Show that Theorem 5.4.1 is applicable if $A = \begin{bmatrix} 1 & 1-i \\ 1+i & 0 \end{bmatrix}$ and verify the conclusion of the Theorem.

SOLUTION. First notice that

$$A^H = \begin{bmatrix} 1 & 1-i \\ 1+i & 0 \end{bmatrix}^H = \begin{bmatrix} 1 & 1+i \\ 1-i & 0 \end{bmatrix}^T = \begin{bmatrix} 1 & 1-i \\ 1+i & 0 \end{bmatrix} = A$$

It follows that A is Hermitian and the theorem is applicable. Now we compute the eigenvalues of A by solving the characteristic equation

$$\begin{aligned} 0 = \det(A - \lambda I) &= \det \begin{bmatrix} 1-\lambda & 1-i \\ 1+i & -\lambda \end{bmatrix} \\ &= (1-\lambda)(-\lambda) - (1+i)(1-i) \\ &= \lambda^2 - \lambda - 2 \\ &= (\lambda+1)(\lambda-2) \end{aligned}$$

Hence the eigenvalues of A are $\lambda = -1, 2$ which are real. □

Caution: Although the *eigenvalues* of a Hermitian matrix are guaranteed to be real, the *eigenvectors* may not be real unless the matrix in question is real.

The Principal Axes Theorem

A key fact about Hermitian matrices is the so-called *principal axes theorem*; its proof is a simple consequence of the Schur triangularization theorem which is proved in Section 5.5. We will content ourselves here with stating the theorem and supplying a proof for the case that the eigenvalues of A are distinct. This proof also shows us one way to carry out the diagonalization process.

Principal Axes Theorem

THEOREM 5.4.3. *Every Hermitian matrix is unitarily diagonalizable, and every real Hermitian matrix is orthogonally diagonalizable.*

PROOF. Let us assume that the eigenvalues of the $n \times n$ matrix A are distinct. We saw in Theorem 5.4.1 of Chapter 4 that the eigenvalues of A are real. Let these eigenvalues be $\lambda_1, \lambda_2, \ldots, \lambda_n$. Now find an eigenvector $\mathbf{v}_k$ for each eigenvalue λ_k. We can assume that each $\mathbf{v}_k$ is unit length by replacing it by the vector divided by its length if necessary. We now have a diagonalizing matrix, as prescribed by the Diagonalization Theorem, namely the matrix $P = [\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n]$.

Remembering that $A\mathbf{v}_j = \lambda_j\mathbf{v}_j$, $A\mathbf{v}_k = \lambda_k\mathbf{v}_k$ and that $A^H = A$, we see that

$$\lambda_k\mathbf{v}_j^H\mathbf{v}_k = \mathbf{v}_j^H\lambda_k\mathbf{v}_k = \mathbf{v}_j^H A\mathbf{v}_k = (A\mathbf{v}_j)^H\mathbf{v}_k = (\lambda_j\mathbf{v}_j)^H\mathbf{v}_k = \lambda_j\mathbf{v}_j^H\mathbf{v}_k$$

Now bring both terms to one side of the equation and factor out the term $\mathbf{v}_j^H\mathbf{v}_k$ to obtain

$$(\lambda_k - \lambda_j)\mathbf{v}_j^H\mathbf{v}_k = 0$$

Thus if $\lambda_k \neq \lambda_j$, it follows that $\mathbf{v}_j \cdot \mathbf{v}_k = \mathbf{v}_j^H\mathbf{v}_k = 0$. In other words the eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ form an orthonormal set. Therefore, the matrix P is unitary. If A is real, then so are the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ and P is orthogonal in this case. □

The proof we have just given suggests a practical procedure for diagonalizing a Hermitian or real symmetric matrix. The only additional information that we need for the complete procedure is advice on what to do if the eigenvalue λ is repeated. This is a sticky point.What we need to do in this case is find an orthogonal basis of the eigenspace $\mathcal{E}_\lambda(A) = \mathcal{N}(A - \lambda I)$. It is always possible to find such a basis using the so-called Gram-Schmidt algorithm, which is discussed in Chapter 6 or the modified Gram-Schmidt algorithm discussed on Page 208. For the hand calculations that we do in this chapter, the worst situation that we will encounter is that the eigenspace $\mathcal{E}_\lambda$ is two-dimensional, say with a basis $\mathbf{v}_1, \mathbf{v}_2$. In this case replace $\mathbf{v}_2$ by $\mathbf{v}_2^* = \mathbf{v}_2 - \mathrm{proj}_{\mathbf{v}_1}\mathbf{v}_2$. We know that $\mathbf{v}_2^*$ is orthogonal to $\mathbf{v}_1$ (see Theorem 6.2.16) so that $\mathbf{v}_1, \mathbf{v}_2^*$is an orthogonal basis of $\mathcal{E}_\lambda(A)$.. We illustrate the procedure with a few examples.

EXAMPLE 5.4.4. Find an eigensystem for the matrix $A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}$ and use this to orthogonally diagonalize A.

SOLUTION. Notice that A is real symmetric, so diagonalizable by the principal axes theorem. First calculate the characteristic polynomial of A as

$$\begin{aligned} |A - \lambda I| &= \begin{vmatrix} 1-\lambda & 2 & 0 \\ 2 & 4-\lambda & 0 \\ 0 & 0 & 5-\lambda \end{vmatrix} \\ &= ((1-\lambda)(4-\lambda) - 2\cdot 2)(5-\lambda) \\ &= -\lambda(\lambda-5)^2 \end{aligned}$$

so that the eigenvalues of A are $\lambda = 0, 5, 5$.

Next find eigenspaces for each eigenvalue. For $\lambda = 0$, we find the null space by row reduction

$$A - 0I = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix} \xrightarrow[E_{21}(-2)]{} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 5 \end{bmatrix} \xrightarrow[\substack{E_{23} \\ E_2(\frac{1}{5})}]{} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

so that the null space is spanned by the vector $(-2, 1, 0)$. Normalize this vector to obtain $\mathbf{v}_1 = (-2, 1, 0)/\sqrt{5}$. Next compute the eigenspace for $\lambda = 5$ via row reductions

$$A - 5I = \begin{bmatrix} -4 & 2 & 0 \\ 2 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[E_{21}(1/2)]{} \begin{bmatrix} -4 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[E_1(-1/4)]{} \begin{bmatrix} 1 & -1/2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which gives two eigenvectors, $(1/2, 1, 0)$ and $(0, 0, 1)$. Normalize these to get $\mathbf{v}_2 = (1, 2, 0)/\sqrt{5}$ and $\mathbf{v}_3 = (0, 0, 1)$. In this case $\mathbf{v}_2$ and $\mathbf{v}_3$ are already orthogonal, so the diagonalizing matrix can be written as

$$P = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = \frac{1}{\sqrt{5}} \begin{bmatrix} -2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

In fact, we can check that

$$P^T A P = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

We leave this calculation to the reader.

PROOF. EXAMPLE 5.4.5. Let $A = \begin{bmatrix} 1 & 1-i \\ 1+i & 0 \end{bmatrix}$ as in Example 5.4.2. Unitarily diagonalize this matrix.

SOLUTION. In Example 5.4.2 we computed the eigenvalues to be $\lambda = -1, 2$. Next find eigenspaces for each eigenvalue. For $\lambda = -1$, we find the null space by row reduction

$$A + I = \begin{bmatrix} 2 & 1-i \\ 1+i & 1 \end{bmatrix} \xrightarrow[E_{21}(-(1+i)/2)]{} \begin{bmatrix} 2 & 1-i \\ 0 & 0 \end{bmatrix} \xrightarrow[E_1(1/2)]{} \begin{bmatrix} 1 & (1-i)/2 \\ 0 & 0 \end{bmatrix}$$

so that the null space is spanned by the vector $((-1+i)/2, 1)$. A similar calculation shows that a basis of eigenvectors for $\lambda = 2$ consists of the vector $(-1, (-1-i)/2)$. Normalize these vectors to obtain $\mathbf{u}_1 = ((-1+i)/2, 1)/\sqrt{3/2}$ and $\mathbf{u}_2 = (-1, (-1-i)/2)/\sqrt{3/2}$. So set

$$U = \sqrt{\frac{2}{3}} \begin{bmatrix} \frac{-1+i}{2} & -1 \\ 1 & \frac{-1-i}{2} \end{bmatrix}$$

and obtain that

$$U^{-1} A U = U^H A U = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}$$

The last calculation is left to the reader. □

5.4 Exercises

1. Show the following matrices are Hermitian and find their eigenvalues:

$$\text{(a)} \begin{bmatrix} 3 & -i \\ i & 1 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} 1 & 1+i & 0 \\ 1-i & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$\text{(d)} \begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix} \quad \text{(e)} \begin{bmatrix} 3 & i \\ -i & 0 \end{bmatrix} \quad \text{(f)} \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & -2 \\ 0 & -2 & 1 \end{bmatrix}$$

2. Find eigensystems for the matrices of Exercise 1 and orthogonal (unitary) matrices that diagonalize these matrices.

3. Show that these matrices are orthogonal and compute their eigenvalues. Determine if it is possible to orthogonally diagonalize these matrices

$$\text{(a)} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \qquad \text{(b)} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad \text{(c)} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$$

4. Let $A = \begin{bmatrix} 3 & i \\ -i & 3 \end{bmatrix}$.

(a) Show that A is a Hermitian matrix.

(b) Find the eigenvalues and eigenvectors of A.

(c) Find a unitary matrix P and diagonal D such that $P^{-1}AP = D$.

(d) Use (c) to find a formula for the kth power of A.

5. Suppose that A is symmetric and orthogonal. Prove that the only possible eigenvalues of A are ± 1.

6. Let A be real symmetric positive definite matrix. Show that A has a real symmetric positive definite square root, that is, there is a symmetric positive definite matrix S such that $S^2 = A$. *Hint:* First show it for a diagonal matrix with positive diagonal entries. Then use Exercise 15 and the principal axes theorem.

7. Let $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ so that the eigenvalues of A are $1, 2$ and 4 (assume this). Use the method of Exercise 6 to find a square root of A.

5.5. *Schur Form and Applications

Recall that matrices A and B are similar if there is an invertible matrix S such that $B = S^{-1}AS$; if the transformation matrix S is unitary, then $S^{-1} = S^H$. The main object of this section is to prove a famous theorem in linear algebra which provides a nice answer to the following question: if we only wish to use orthogonal (or unitary) matrices as similarity transformation matrices, what is the simplest form to which a given matrix A can be transformed. It would be nice if we could say something like "diagonal" or "Jordan canonical form." Unfortunately, neither is possible. However, upper triangular matrices are very nice special forms of matrices. In particular, we can see the eigenvalues of an upper triangular matrix at a glance. That makes the following theorem extremely attractive. Its proof is also very interesting, in that it actually suggests an algorithm for computing the so-called Schur triangular form.

THEOREM 5.5.1. *Let A be an arbitrary square matrix. Then there exists a unitary matrix U such that $U^T AU$ is an upper triangular matrix. If A and its eigenvalues are real, then U can be chosen to be orthogonal.* Schur Triangularization Theorem

PROOF. We will show how to triangularize A one column at time. First we show how to start the process. Compute an eigenvalue λ_1 of A and a corresponding eigenvector $\mathbf{w}$ of unit length in the standard norm. We may assume that the first coordinate of $\mathbf{w}$ is real. If not, replace $\mathbf{w}$ by $e^{-i\theta}\mathbf{w}$ where θ is a polar argument of the first coordinate of $\mathbf{w}$. This does not affect the length of $\mathbf{w}$ and any multiple of $\mathbf{w}$ is still an eigenvector of A. Now let $\mathbf{v} = \mathbf{w} - \mathbf{e}_1$, where $\mathbf{e}_1 = (1, 0, \ldots, 0)$. We make the convention that $H_0 = I$. Form the (possibly complex) Householder matrix $H_{\mathbf{v}}$. Since $\mathbf{w} \cdot \mathbf{e}_1$ is real, it follows from Exercise 3 that $H_{\mathbf{v}}\mathbf{w} = \mathbf{e}_1$. Now recall that Householder matrices are unitary and symmetric, so that $H_{\mathbf{v}}^H = H_{\mathbf{v}} = H_{\mathbf{v}}^{-1}$. Hence

$$H_{\mathbf{v}}^H A H_{\mathbf{v}} \mathbf{e}_1 = H_{\mathbf{v}} A H_{\mathbf{v}}^{-1} \mathbf{e}_1 = H_{\mathbf{v}} A \mathbf{w} = H_{\mathbf{v}} \lambda_1 \mathbf{w} = \lambda_1 \mathbf{e}_1$$

Therefore, the entries under the first row and in the first column of $H_{\mathbf{v}}^H A H_{\mathbf{v}}$ are zero.

Suppose we have reached the kth stage ($k = 0$ is start) where we have a unitary matrix V_k such that

$$V_k^H A V_k = \begin{bmatrix} \lambda_1 & * & \cdots & * \\ \vdots & \ddots & * & \vdots \\ 0 & \cdots & \lambda_k & * \\ 0 & \cdots & 0 & B \end{bmatrix} = \begin{bmatrix} R_k & C \\ 0 & B \end{bmatrix}$$

with the submatrix R_k upper triangular. Compute an eigenvalue λ_{k+1} of the submatrix B and a corresponding eigenvector $\mathbf{w}$ of unit length in the standard norm. Now repeat the argument of the first paragraph with B in place of A to obtain a Householder matrix $H_{\mathbf{v}}$ of the same size as B such that the entries under the first row and in the first column of $H_{\mathbf{v}}^H B H_{\mathbf{v}}$ are zero. Form the unitary matrix

$$V_{k+1} = \begin{bmatrix} I_k & 0 \\ 0 & H_{\mathbf{v}} \end{bmatrix} V_k$$

and obtain that

$$\begin{aligned} V_{k+1}^H A V_{k+1} &= \begin{bmatrix} I_k & 0 \\ 0 & H_{\mathbf{v}} \end{bmatrix} V_k^H A V_k \begin{bmatrix} I_k & 0 \\ 0 & H_{\mathbf{v}} \end{bmatrix} \\ &= \begin{bmatrix} I_k & 0 \\ 0 & H_{\mathbf{v}} \end{bmatrix} \begin{bmatrix} R_k & C \\ 0 & B \end{bmatrix} \begin{bmatrix} I_k & 0 \\ 0 & H_{\mathbf{v}} \end{bmatrix} \\ &= \begin{bmatrix} R_k & C H_{\mathbf{v}} \\ 0 & H_{\mathbf{v}}^H B H_{\mathbf{v}} \end{bmatrix} \end{aligned}$$

This new matrix is upper triangular in the first $k + 1$ columns, so we can continue in this fashion until we reach the last column, at which point we set $U = V_n$ to obtain that $U^H AU$ is upper triangular. Finally, notice that if the eigenvalues and eigenvectors that we calculate are real, which would certainly be the case if A and the eigenvalues of A were real, then the Householder matrices used in the proof are all real, so that the matrix U is orthogonal. □

Of course, the upper triangular matrix T and triangularizing matrix U are not unique. Nonetheless, this is a very powerful theorem. Consider what it says in the case that A is Hermitian: the Principal Axes Theorem is a simple special case of it.

COROLLARY 5.5.2. *Every Hermitian matrix is unitarily (orthogonally, if the matrix is real) diagonalizable.* Principal Axes Theorem

PROOF. Let A be Hermitian. According to the Schur triangularization theorem there is a unitary matrix U such that $U^H AU = R$ is upper triangular. We check that

$$R^H = \left(U^H AU\right)^H = U^H A^H \left(U^H\right)^H = U^H AU = R.$$

Therefore R is both upper and lower triangular. This makes R a diagonal matrix and proves the theorem, except for the fact that U may be chosen orthogonal if A is real. To see this last fact, notice that if A is real symmetric, then A and its eigenvalues are real, so according to the triangularization theorem, U can be chosen orthogonal. □

As a last application of the Schur triangularization theorem, we show the real significance of *normal* matrices. This term has appeared in several exercises. Recall that a (square) matrix A is normal if $A^H A = AA^H$.

COROLLARY 5.5.3. *A matrix is unitarily diagonalizable if and only if it is normal.*

PROOF. It is easy to see that every unitarily diagonalizable matrix is normal. We leave this as an exercise.

Let A be normal. According to the Schur triangularization theorem there is a unitary matrix U such that $U^H AU = R$ is upper triangular. But then we have that $R^H = U^H A^H U$, so that

$$R^H R = U^H A^H U U^H AU = U^H A^H AU = U^H AA^H U = U^H AU U^H A^H U = RR^H$$

Therefore R commutes with R^H, which means that R is diagonal by Exercise . This completes the proof. □

Our last application of Schur's theorem is a far reaching generalization of Theorem 5.1.5.

COROLLARY 5.5.4. *Let $f(x)$ and $g(x)$ be polynomials and A a square matrix such that $g(A)$ is invertible (e.g., g(x)=1). Then the eigenvalues of the matrix $f(A)g(A)^{-1}$ are of the form $f(\lambda)/g(\lambda)$, where λ runs over the eigenvalues of A.*

PROOF. We sketch the proof. As a first step, we make two observations about upper triangular matrices S and T with diagonal terms $\lambda_1, \lambda_2, \ldots, \lambda_n$, and $\mu_1, \mu_2, \ldots, \mu_n$, respectively.

1. ST is upper triangular with diagonal terms $\lambda_1\mu_1, \lambda_2\mu_2, \ldots, \lambda_n\mu_n$.
2. If S is invertible, then S is upper triangular with diagonal terms $1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_n$.

Now we have seen in Exercise 12 of Section 5.2 that for any invertible P of the right size, $P^{-1}f(A)P = f(P^{-1}AP)$. Similarly, if we multiply the identity $g(A)g(A)^{-1} = I$ by P^{-1} and P, we see that $P^{-1}g(A)^{-1}P = g(P^{-1}AP)^{-1}$. Thus, if P is a matrix that unitarily diagonalizes A, then

$$P^{-1}f(A)g(A)^{-1}P = f(P^{-1}AP)g(P^{-1}AP)^{-1}$$

so that by our first observations, this matrix is upper triangular with diagonal entries of the required form. Since similar matrices have the same eigenvalues, it follows that the eigenvalues of $f(A)g(A)^{-1}$ are of the required form. □

5.5 Exercises

1. The matrix

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

has 3 as an eigenvalue. Carry out the first step of Schur triangulation on A.

2. Prove that every unitarily diagonalizable matrix is normal.

3. Use Corollary 5.5.2 to show that the eigenvalues of a Hermitian matrix must be real.

4. Prove that if an upper triangular matrix commutes with its Hermitian transpose, then the matrix must be diagonal. *Hint:* Equate $(1, 1)$th coefficients of the equation $R^H R = RR^H$ and see what can be gained from it. Proceed to the $(2, 2)$th coefficient, etc.

5. Show that if $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, $\mathbf{x}$ and $\mathbf{y}$ have the same length and $\mathbf{x} \cdot \mathbf{y}$ is real, then $\mathbf{x} + \mathbf{y}$ is orthogonal to $\mathbf{v} = \mathbf{x} - \mathbf{y}$.

6. With $\mathbf{x}, \mathbf{y}, \mathbf{v}$ as in Exercise 5, show that

$$\mathbf{x} = \frac{1}{2}(\mathbf{x} - \mathbf{y}) + (\mathbf{x} + \mathbf{y}) = \mathbf{p} + \mathbf{u}$$

where $\mathbf{p}$ is parallel to $\mathbf{v}$ and $\mathbf{u}$ is orthogonal to $\mathbf{v}$.

7. With $\mathbf{x}, \mathbf{y}, \mathbf{v}$ as in Exercise 5, show that $H_v\mathbf{x} = \mathbf{y}$. *Hint:* A text theorem about Householder matrices applies to this setup.

5.6. *The Singular Value Decomposition

The object of this section is to develop yet one more factorization of a matrix that tells us a lot about the matrix. For simplicity, we stick with the case of a real matrix A and orthogonal matrices. However, the factorization we are going to discuss can be done with complex A and unitary matrices. This factorization is called the *singular value decomposition (SVD for short)*. It has a long history in matrix theory, but was popularized in the sixties as a powerful computational tool. Here is the basic question that it answers: if multiplication on one side can produce an upper triangular matrix, as in the QR factorization, how simple a matrix can be produced by multiplying on each side by a (possibly different) orthogonal matrix? The answer, as you might guess, is a matrix that is both upper and lower triangular, that is, diagonal. However, verification of this fact is much more subtle than that of one sided factorizations such as the QR factorization. Here is the key result:

THEOREM 5.6.1. *Let A be an $m \times n$ real matrix. Then there exist $m \times m$ orthogonal matrix U, $n \times n$ orthogonal matrix V and $m \times n$ diagonal matrix Σ with diagonal entries $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p$, with $p = \min\{m, n\}$, such that $U^T AV = \Sigma$. Moreover, the numbers $\sigma_1, \sigma_2, \ldots, \sigma_p$ are uniquely determined by A.*

Singular Value Decomposition

PROOF. There is no loss of generality in assuming that $n = \min\{m, n\}$. For if this is not the case, we can prove the theorem for A^T and by transposing the resulting SVD for A^T, obtain a factorization for A. Form the $n \times n$ matrix $B = A^T A$. This matrix is symmetric and its eigenvalues are nonnegative (we leave these facts as exercises). Because they are nonnegative, we can write the eigenvalues of B in decreasing order of magnitude as the squares of positive real numbers, say as $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_n^2$. Now we know from the Principal Axes Theorem that we can find an orthonormal set of eigenvectors corresponding to these eigenvalues, say $B\mathbf{v}_k = \sigma_k^2 \mathbf{v}_k, \quad k = 1, 2, \ldots, n$. Let $V = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$. Then V is an orthogonal $n \times n$ matrix. Next, suppose that $\sigma_{r+1}, \sigma_{r+2}, \ldots, \sigma_n$ are zero, while $\sigma_r \neq 0$.

Next set $\mathbf{u}_j = \frac{1}{\sigma_j} A\mathbf{v}_j, \ j = 1, 2, \ldots, r$. These are orthonormal vectors in $\mathbb{R}^m$ since

$$\mathbf{u}_j^T \mathbf{u}_k = \frac{1}{\sigma_j \sigma_k} \mathbf{v}_j^T A^T A \mathbf{v}_k = \frac{1}{\sigma_j \sigma_k} \mathbf{v}_j^T B \mathbf{v}_k = \frac{\sigma_k^2}{\sigma_j \sigma_k} \mathbf{v}_j^T \mathbf{v}_k = \begin{cases} 0, \text{if} j \neq k \\ 1, \text{if} j = k \end{cases}$$

Now expand this set to an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ of $\mathbb{R}^m$. This is possible by Theorem 4.3.12 in Section 4.3. Now set $U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m]$. This matrix is orthogonal and we calculate that if $k > r$, then $\mathbf{u}_j^T A\mathbf{v}_k = 0$ since $A\mathbf{v}_k = 0$, and if $k < r$, then

$$\mathbf{u}_j^T A\mathbf{v}_k = \mathbf{u}_j^T A\mathbf{v}_k = \sigma_k \mathbf{u}_j^T \mathbf{u}_k = \begin{cases} 0, & \text{if } j \neq k \\ \sigma_k, & \text{if } j = k \end{cases}$$

It follows that $U^T AV = [\mathbf{u}_j^T A\mathbf{v}_k] = \Sigma$.

Finally, if U, V are orthogonal matrices such that $U^T AV = \Sigma$, then $A = U\Sigma V^T$ and therefore

$$B = A^T A = V\Sigma U^T U\Sigma V^T = V\Sigma^2 V^T$$

so that the squares of the diagonal entries of Σ are the eigenvalues of B. It follows that the numbers $\sigma_1, \sigma_2, \ldots, \sigma_n$ are uniquely determined by A. □

NOTATION 5.6.2. The numbers $\sigma_1, \sigma_2, \ldots, \sigma_p$ are called the *singular values* of the matrix A, the columns of U are the *left singular vectors* of A, and the columns of V are the *right singular values* of A.

There is an interesting geometrical interpretation of this theorem from the perspective of linear transformations and change of basis as developed in Section 3.7. It can can be stated as follows.

COROLLARY 5.6.3. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with matrix A with respect to the standard bases. Then there exist orthonormal bases $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, such that the matrix of T with these bases is diagonal with nonnegative entries down the diagonal.*

PROOF. First observe that if $U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m]$ and $V = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$, then U and V are the change of basis matrices from the standard bases to the bases $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. Also, $U^{-1} = U^T$. Now apply Corollary 3.7.5 of Section 3.7 and the result follows. □

We leave the following fact as an exercise.

COROLLARY 5.6.4. *Let $U^T AV = \Sigma$ be the SVD of A and suppose that $\sigma_r \neq 0$ and $\sigma_{r+1} = 0$. Then*

1. $\operatorname{rank} A = r$
2. $\ker A = \operatorname{span}\{\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n\}$
3. $\operatorname{range} A = \operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$

We have only scratched the surface of the many facets of the SVD. Like most good ideas, it is rich in applications. We mention one more. It is based on the following fact, which can be proved by examining the entries of $A = U\Sigma V^T$: The matrix A of rank r can be written as a sum of r rank one matrices, namely

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

where the $\sigma_k, \mathbf{u}_k, \mathbf{v}_k$ are the singular values, left and right singular vectors, respectively. In fact, it can be shown that this representation is the most economical in the sense that the partial sums

$$\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T, \qquad k = 1, 2, \ldots, r$$

give the rank k approximation to A that is closest among all rank k approximations to A. This gives us an intriguing way to compress data in a lossy way (i.e., with some loss of data). For example, suppose A is a matrix of floating point numbers representing a picture. We might get a reasonable good approximation to the picture by using only the σ_k larger than a certain threshhold. Thus, with a $1,000 \times 1,000$, matrix A that has a very small σ_{21}, we could get by with the data $\sigma_k, \mathbf{u}_k, \mathbf{v}_k$, $k = 1, 2, \ldots, 20$. Consequently, we would only store these quantities, which add up to $1000 \times 40 + 20 = 40,020$ numbers. Contrast this with storing the full matrix of $1,000 \times 1,000 = 1,000,000$ entries, and you can see the gain in economy.

5.6 Exercises

1. Exhibit a singular value decomposition for the following matrices.

$$\text{(a)} \begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \text{(b)} \begin{bmatrix} -2 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \text{(c)} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

2. Express the singular values and vectors of A^T in terms of those of A.

3. Trace through the proof of the SVD with the matrix and construct the SVD of the matrix

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

4. Find the least squares solution to the problem $Ax = b$ when A only has nonzero entries along the main diagonal. Then use this solution to design an algorithm for solving the general least squares problem by means of the SVD of A.

5. Let A be a real matrix and U, V orthogonal matrices.

(a) Show from definition that $\|U^T AV\|_2 = \|A\|_2$

(b) Determine $\|\Sigma\|_2$ if Σ is a diagonal matrix with non-negative entries.

(c) Use parts (a),(b) and the SVD to express $\|A\|_2$ in terms of the singular values of A.

6. Prove Corollary 5.6.4.

7. Digitize a picture into a 640×400 (standard VGA) matrix of greyscale pixels, where the value of each pixel is a number x, $0 \leq x \leq 1$, with black corresponding to $x = 0$ and white to $x = 1$. Compute the SVD of this image matrix and display various approximations using $10, 20$ and 40 of the singular values and vector pairs. Do any of these give a good visual approximation to the picture? If not, find a minimal number that works. You will need computer support for this exercise.

5.7. *Computational Notes and Projects

Computation of Eigensystems

Nowadays, one can use a MAS like MATLAB or Octave on a home PC to find a complete eigensystem for, say a 100×100 matrix, in less than a second. That's pretty remarkable and, to some extent, a tribute to the fast cheap hardware commonly available to the public. But hardware is only part of the story. Bad computational algorithms can bring the fastest computer to its knees. The rest of the story concerns the remarkable developments in numerical linear algebra over the past 50 years which have given us fast reliable algorithms for eigensystem calculation. We can only scratch the surface of these developments in this brief discussion. At the outset, we rule out the methods developed in this chapter as embodied in the eigensystem algorithm (Page 216). These are for simple hand calculations and theoretical purposes. See the polynomial equations project that follows this discussion for some more comments about root finding.

We are going to examine some iterative methods for selectively finding eigenpairs of a real matrix whose eigenvalues are real and distinct. Hence the matrix A is diagonalizable. The hypothesis of diagonalizability may seem too constraining, but there is this curious aphorism that "numerically every matrix is diagonalizable." The reason is as follows: once you perform store and numerical calculations on the entries of A, you perturb them a small essentially random amount. This has the effect of perturbing the eigenvalues of the calculated A a small random amount. Thus, the probability that any two eigenvalues of A are numerically equal is quite small. To focus matters, consider the test matrix

$$A = \begin{bmatrix} -8 & -5 & 8 \\ 6 & 3 & -8 \\ -3 & 1 & 9 \end{bmatrix}$$

Just for the record, the actual eigenvalues of A are $-2, 1$ and 5 (see Exercise 1 for an explanation). Now we ask three questions about A :

(1) How can we get a ballpark estimate of the location of the eigenvalues of A?

(2) How can we estimate the *dominant* eigenpair $(\lambda, \mathbf{x})$ of A? (*Dominant* means that λ is larger in absolute value than any other eigenvalue of A.

(3) Given a good estimate of any eigenvalue λ of A, how can we improve the estimate and compute a corresponding eigenvector?

One answer to question (1) is the following theorem, which predates modern numerical analysis, but has proved to be quite useful. Because it helps locate eigenvalues, it is called a "localization theorem."

Gershgorin Circle Theorem

THEOREM 5.7.1. *Let $A = [a_{ij}]$ be an $n \times n$ matrix and define disks in the complex plane by*

$$\begin{aligned} r_j &= \sum_{\substack{k=1 \\ k \neq j}}^{n} |a_{jk}| \\ C_j &= \{\, z \mid |z - r_j| \leq |a_{jj}| \end{aligned}$$

Then

1. Every eigenvalue of A is contained in some disk C_j.
2. If k of the disks are disjoint from the others, then exactly k eigenvalues are contained in the union of these disks.

PROOF. To prove 1, let λ be an eigenvalue of A and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be an eigenvector corresponding to λ. Suppose that x_j is the largest coordinate of $\mathbf{x}$ in absolute value. Divide $\mathbf{x}$ by this entry to obtain an eigenvector whose largest coordinate is $x_j = 1$. Without loss of generality, this vector is $\mathbf{x}$. Consider the jth entry of the zero vector $\lambda\mathbf{x} - A\mathbf{x}$ which is

$$(\lambda - a_j)1 + \sum_{\substack{k=1 \\ k \neq j}}^{n} a_{jk} x_k = 0.$$

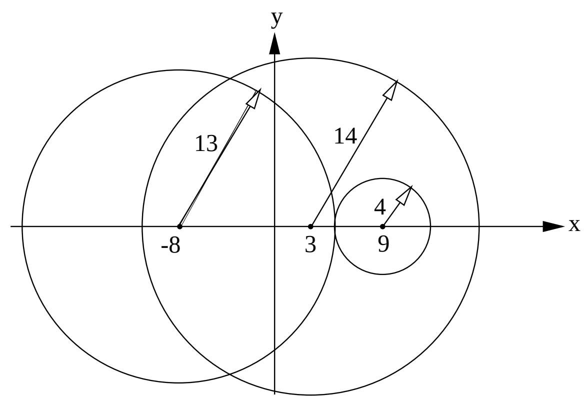


FIGURE 5.7.1. Gershgorin circles for A.

Bring the sum to the right hand side and take absolute values to obtain

$$\begin{aligned} |\lambda - a_j| &= \Big| \sum_{\substack{k=1 \\ k \neq j}}^{n} a_{jk} x_k \Big| \\ &\leq \sum_{\substack{k=1 \\ k \neq j}}^{n} |\, a_{jk} \,|\, |\, x_k \,| \leq r_j \end{aligned}$$

since each $|\, x_k \,| \leq 1$. This shows that $\lambda \in C_j$ which proves 1. We will not prove 2, as it requires some complex analysis. □

EXAMPLE 5.7.2. Apply the Gershgorin circle theorem to A and sketch the resulting Gershgorin disks.

SOLUTION. The circles are easily seen to be

$$\begin{aligned} C_1 &= \{\, z \,|\, |z+8| \leq 13\} \\ C_2 &= \{\, z \,|\, |z-3| \leq 14\} \\ C_3 &= \{\, z \,|\, |z-9| \leq 4\} \end{aligned}$$

A sketch of them is provided in Figure 5.7.1. □

Now we turn to question (2). One answer to it is contained in the following algorithm, known as the *power method*.

Power Method: To compute an approximate eigenpair $(\lambda,\mathbf{x})$ of A with $\|\mathbf{x}\| = 1$ and λ the dominant eigenvalue.

1. Input an initial guess $\mathbf{x}_0$ for $\mathbf{x}$
2. For $k = 0, 1, \ldots$ until convergence of $\lambda^{(k)}$'s:
   (a) $\mathbf{y} = A\mathbf{x}_k$
   (b) $\mathbf{x}_{k+1} = \dfrac{\mathbf{y}}{\|\mathbf{y}\|}$

(c) $\lambda^{(k+1)} = \mathbf{x}_{k+1}^T A \mathbf{x}_{k+1}$

That's all there is to it! Why should this algorithm converge? The secret to this algorithm lies in a formula we saw earlier in our study of discrete dynamical systems, namely Equation 5.3.2 which we reproduce here

$$\mathbf{x}^{(k)} = A^k \mathbf{x}^{(0)} = c_1 \lambda_1^k \mathbf{v}_1 + c_2 \lambda_2^k \mathbf{v}_2 + \cdots + c_n \lambda_n^k \mathbf{v}_n.$$

Here it is understood that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis of eigenvectors corresponding to eigenvalues $\lambda_1 \lambda_2, \ldots, \lambda_n$, which, with no loss of generality, we can assume to be unit length vectors. Notice that at each stage of the power method we divided the computed iterate $\mathbf{y}$ by its length to get the next $\mathbf{x_{k+1}}$, and this division causes no directional change. Thus we would get exactly the same vector if we simply set $\mathbf{x}_{k+1} = \mathbf{x}^{(k+1)} / \left\| \mathbf{x}^{(k+1)} \right\|$. Now for large k the ratios $(\lambda_j / \lambda_1)^k$ can be made as small as we please, so we can rewrite the above equation as

$$\mathbf{x}^{(k)} = A^k \mathbf{x}^{(0)} = \lambda_1^k \left\{ c_1 \mathbf{v}_1 + c_2 \left(\frac{\lambda_2}{\lambda_1} \right)^k \mathbf{v}_2 + \cdots + c_n \left(\frac{\lambda_n}{\lambda_1} \right)^k \mathbf{v}_n \right\} \approx \lambda_1^k c_1 \mathbf{v}_1.$$

Assuming that $c_1 \neq 0$, which is likely if $\mathbf{x}_0$ is randomly chosen, we see that

$$\begin{aligned} \mathbf{x}_{k+1} &= \frac{A\mathbf{x}^{(k)}}{\left\| A\mathbf{x}^{(k)} \right\|} \approx \frac{\lambda_1^k c_1 \lambda_1 \mathbf{v}_1}{\left| \lambda_1^k c_1 \lambda_1 \right|} = \pm \mathbf{v}_1 \\ \lambda^{(k+1)} &= \mathbf{x}_{k+1}^T A \mathbf{x}_{k+1} \approx (\pm \mathbf{v}_1)^T A (\pm \mathbf{v}_1) = \lambda_1 \end{aligned}$$

Thus we see that the sequence of $\lambda^{(k)}$'s converges to λ_1 and the sequence of $\mathbf{x}_k$'s converges to $\pm \mathbf{v}_1$. The argument (it isn't rigorous enough to be called a proof) we have just given shows that the oscillation in sign in $\mathbf{x}_k$occurs in the case $\lambda < 0$. You might notice also that the argument doesn't require the initial guess to be a real vector. Complex vectors are permissible.

If we apply the power method to our test problem with an initial guess of $\mathbf{x}_0 = (1, 1, 1)$, we get every third value as follows:

k	$\lambda^{(k)}$	$\mathbf{x}_k$
0		$(1, 1, 1)$
3	5.7311	$(0.54707, -0.57451, 0.60881)$
6	4.9625	$(0.57890, -0.57733, 0.57581)$
9	5.0025	$(0.57725, -0.57735, 0.57745)$
12	4.9998	$(0.57736, -0.57735, 0.57734)$

Notice, that the eigenvector looks a lot like a multiple of $(1, -1, 1)$ and the eigenvalue looks a lot like 5. This is an exact eigenpair, as one can check.

Finally, we turn to question (3). One answer to it is contained in the following algorithm, known as the *inverse iteration method*.

Inverse Iteration Method: To compute an approximate eigenpair $(\lambda,\mathbf{x})$ of A with $\|\mathbf{x}\| = 1$.

1. Input an initial guess $\mathbf{x}_0$ for $\mathbf{x}$ and a *close* approximation $\mu = \lambda_0$ to λ.
2. For $k = 0, 1, \ldots$ until convergence of $\lambda^{(k)}$'s:
   (a) $\mathbf{y} = (A - \mu I)^{-1} \mathbf{x}_k$

(b) $\mathbf{x}_{k+1} = \dfrac{\mathbf{y}}{\|\mathbf{y}\|}$

(c) $\lambda^{(k+1)} = \mathbf{x}_{k+1}^T A \mathbf{x}_{k+1}$

Notice that the inverse iteration method is simply the power method applied to the matrix $(A - \mu I)^{-1}$. In fact, it is sometimes called the inverse power method. The scalar μ is called a *shift*. Here is the secret of success for this method: we assume that μ is closer to a definite eigenvalue λ of A than to any other eigenvalue. But we don't want too much accuracy! We need $\mu \neq \lambda$. Theorem 5.1.5 in Section 1 of this chapter shows that the eigenvalues of the matrix $A - \mu I$ are of the form $\sigma - \mu$ where σ runs over the eigenvalues of A. Thus the matrix $A - \mu I$ is nonsingular since no eigenvalue is zero, and Exercise 11 shows us that the eigenvalues of $(A - \mu I)^{-1}$ are of the form $1/(\sigma - \mu)$ where σ runs over the eigenvalues of A. Since μ is closer to λ than to any other eigenvalue of A, the eigenvalue $1/(\lambda - \mu)$ is the dominant eigenvalue of $(A - \mu I)^{-1}$, which is exactly what we need to make the power method work on $(A - \mu I)^{-1}$. Indeed, if μ is *very* close (but not equal!) to λ convergence should be very rapid.

In a general situation, we could now have the Gershgorin circle theorem team up with inverse iteration. Gershgorin would put us in the right ballpark for values of μ and inverse iteration would finish the job. Let's try this with our test matrix and choices of μ in the interval $[-21, 17]$ suggested by Gershgorin. Let's try $\mu = 0$. Here are the results in tabular form.

k	$\lambda^{(k)}$	$\mathbf{x}_k$ with $\mu = 0.0$
0	0.0	$(1, 1, 1)$
3	0.77344	$(-0.67759, 0.65817, -0.32815)$
6	1.0288	$(-0.66521, 0.66784, -0.33391)$
9	0.99642	$(-0.66685, 0.66652, -0.33326)$
12	1.0004	$(-0.66664, 0.66668, -0.33334)$

It appears that inverse iteration is converging to $\lambda = 1$ and the eigenvector looks suspiciously like a multiple of $(-2, 2, -1)$. This is in fact an exact eigenpair.

There is much more to modern eigenvalue algorithms than we have indicated here. Central topics include deflation, the QR algorithm, numerical stability analysis and many other issues. The interested reader might consult more advanced text such as references such as [**5**], [**4**], [**8**] or [**3**], to name a few.

Project Topics

Project: Solving Polynomial Equations

In homework problems we solve for the roots of the characteristic polynomial in order to get eigenvalues. To this end we can use algebra methods or even Newton's method for numerical approximations to the roots. This is the conventional wisdom usually proposed in introductory linear algebra. But for larger problems than the simple 2×2 or 3×3 matrices we encounter, this method can be too slow and inaccurate. In fact, numerical methods hiding under the hood in a MAS (and some CASs) for finding eigenvalues are so efficient that it is better to turn this whole procedure on its head. Rather than find roots to solve linear algebra (eigenvalue) problems, we can use (numerical) linear algebra to find roots of polynomials. In this project we discuss this methodology and document it in a fairly nontrivial example.

Given a polynomial $f(x) = c_0 + c_1 x + \cdots + c_{n-1}x^{n-1} + x^n$, form the *companion matrix* of $f(x)$

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

It is a key fact that the eigenvalues of $C(f)$ are precisely the roots of the equation $f(x) = 0$. Experiment with $n = 2, 3, 4$ and try to find a proof by expansion across the bottom row of $\det(A - \lambda I)$ that this result is true for all n.

Then use a CAS (or MAS) to illustrate this method by finding approximate roots of three polynomials: a cubic and quartic of your choice and then the polynomial

$$f(x) = 5 + 11x + 4x^2 + 6x^3 + x^4 - 15x^5 + 5x^6 - 3x^7 - 2x^8 + 8x^9 - 5x^{10} + x^{11}$$

In each case use Newton's method to improve the values of some of the roots (it works with complex numbers as well as reals, provided one starts close enough to a root.) Check your answers to this problem by evaluating the polynomial. Use your results to write the polynomial as a product of the linear factors $x - \lambda$, where λ is a root and check the correctness of this factorization.

Project: Finding a Jordan Canonical Form A challenge: Find the Jordan canonical form of this matrix, which is given *exactly* as follows. The solution will require some careful work with a CAS or (preferably) MAS.

$$A = \begin{bmatrix} 1 & 1 & 1 & -2 & 1 & -1 & 2 & -2 & 4 & -3 \\ -1 & 2 & 3 & -4 & 2 & -2 & 4 & -4 & 8 & -6 \\ -1 & 0 & 5 & -5 & 3 & -3 & 6 & -6 & 12 & -9 \\ -1 & 0 & 3 & -4 & 4 & -4 & 8 & -8 & 16 & -12 \\ -1 & 0 & 3 & -6 & 5 & -4 & 10 & -10 & 20 & -15 \\ -1 & 0 & 3 & -6 & 2 & -2 & 12 & -12 & 24 & -18 \\ -1 & 0 & 3 & -6 & 2 & -5 & 15 & -13 & 28 & -21 \\ -1 & 0 & 3 & -6 & 2 & -5 & 15 & -11 & 32 & -24 \\ -1 & 0 & 3 & -6 & 2 & -5 & 15 & -14 & 37 & -26 \\ -1 & 0 & 3 & -6 & 2 & -5 & 15 & -14 & 36 & -25 \end{bmatrix}$$

Your main task is to devise a strategy for identifying the Jordan canonical form matrix J. Do *not* expect to find the invertible matrix S for which $J = S^{-1}AS$. However, a key fact to keep in mind is that if A and B are similar matrices, i.e., $A = S^{-1}BS$ for some invertible S, then $\operatorname{rank} A = \operatorname{rank} B$. In particular, if S is a matrix that puts A into Jordan canonical form, then $J = S^{-1}AS$.

First prove this rank fact for A and B. Show it applies to $A - cI$ and $B - cI$ as well, for any scalar c. Then extend it to powers of A and B.

Now you have the necessary machinery for determining numerically the Jordan canonical form. As a first step, one can use a CAS or MAS to find the eigenvalues of A. Of course, these will only be approximate, so one has to decide how many eigenvalues are really repeated.

Next, one has to determine the number of Jordan blocks of a given type. Suppose λ is an eigenvalue and find the rank of various powers of $A - \lambda I$. It would help greatly in understanding how all this counts blocks if you first experiment with a matrix already in Jordan canonical form, say, for example,

$$J = \begin{bmatrix} J_1(2) & 0 & 0 \\ 0 & J_2(2) & 0 \\ 0 & 0 & J_1(3) \end{bmatrix}.$$

Project: Classification of Quadratic Forms

Recall from calculus that in order to classify all quadratic equations in x and y one went through roughly three steps. First, do a rotation transformation of coordinates to get rid of mixed terms like, say, $2xy$ in the quadratic equation $x^2 + 2xy - y^2 + x - 3y = 4$. Second, do a translation of coordinates to put the equation in a "standard form." Third, identify the curve by your knowledge of the shape of a curve in the given standard form. Standard forms were equations like $x^2/4 + y^2/2 = 1$. Also recall from your calculus study of the conics that it was the second degree terms alone that determined the nature of a quadratic. For example, the second degree terms of the equation above are $x^2 + 2xy - y^2$. The discriminant of the equation was determined by these terms. In this case the discriminant is 8, which tells us that the curve represented by this equation is a hyperbola. Finally, recall that when it came to *quadric* equations, i.e., quadratic equations in 3 unknowns, your text simply provided some examples in "standard form" (six of them to be exact) and maybe mumbled something about this list being essentially all surfaces represented by quadric equations.

Now you are ready for the rest of the story. Just as with curves in x and y, the basic shape of the surface of a quadric equation in x, y and z is determined by the second degree terms. Since this is so, we will focus on an example with no first degree terms, namely,

$$Q(x, y, z) = 2x^2 + 4y^2 + 6z^2 - 4xy - 2xz + 2yz = 1.$$

The problem is simply this: find a change of coordinates that will make it clear which of the six standard forms is represented by this surface. Here is how to proceed: first you must express the so-called "quadratic form" $Q(x, y, z)$ in matrix form as $Q(x, y, z) = [x, y, z]A[x, y, z]^T$. It is easy to find such matrices A. But any such A won't do. Next, you must replace A by the equivalent matrix $(A + A^T)/2$. (Check that if A specifies the quadratic form Q, then so will $(A + A^T)/2$.) The advantage of this latter matrix is that it is symmetric. Now our theory of symmetric matrices can be brought to bear. In particular, we know that there is an orthogonal matrix P such that $P^T AP$ is diagonal, provided A is symmetric. So make the linear change of variables $[x, y, z]^T = P[x', y', z']^T$ and get that $Q(x, y, z) = [x', y', z']P^T AP[x', y', z']^T$. But when the matrix in the middle is diagonal, we end up with squares of x', y' and z', and no mixed terms.

Use the computer algebra system available to you to calculate a symmetric A and to find the eigenvalues of this A. From this data alone you will be able to classify the surface represented by the above equation. Also find unit length eigenvectors for each eigenvalue. Put these together to form the desired orthogonal matrix P which eliminates mixed terms.

An outstanding reference on this topic and many others relating to matrix analysis is the recently republished textbook [**1**] by Richard Bellman which is widely considered to be a classic in the field.

A Report Topic: Management of Sheep Populations

Description of the problem: You are working for the New Zealand Department of Agriculture on a project for sheep farmers. The species of sheep that these shepherds raise have a life-span of 12 years. Of course, some live longer but they are sufficiently few in number and their reproductive rate is so low that they may be ignored in your population study. Accordingly, you divide sheep into 12 age classes, namely those in the first year of life, etc. You have conducted an extensive survey of the demographics of this species of sheep and obtained the following information about the demographic parameters a_i and b_i, where a_i is the reproductive rate for sheep in the ith age class and b_i is the survival rate for sheep in that age class, i.e., the fraction of sheep in that age class that survive to the $i + 1$th class. (As a matter of fact, this table is related to real data. The interested reader might consult the article [**2**] in the bibliography.)

i	1	2	3	4	5	6	7	8	9	10	11	12
a_i	.000	.023	.145	.236	.242	.273	.271	.251	.234	.229	.216	.210
b_i	.845	.975	.965	.950	.926	.895	.850	.786	.691	.561	.370	-

The problem is as follows: in order to maintain a constant population of sheep, shepherds will harvest a certain number of sheep each year. Harvesting need not mean slaughter; it can be accomplished by selling animals to other shepherds, for example. It simply means removing sheep from the population. Denote the fraction of sheep which are removed from the ith age group at the end of each growth period (a year in our case) by h_i. If these numbers are constant from year to year, they constitute a *harvesting policy.* If, moreover, the yield of each harvest, i.e., total number of animals harvested each year, is a constant and the age distribution of the remaining populace is essentially constant after each harvest, then the harvesting policy is called *sustainable.* If all the h_i's are the same, say h, then the harvesting policy is called *uniform*. An advantage of uniform policies is that they are simple to implement: One selects the sheep to be harvested at random.

Your problem: Find a uniform sustainable harvesting policy to recommend to shepherds, and find the resulting distribution of sheep that they can expect with this policy. Shepherds who raise sheep for sale to markets are also interested in a sustainable policy that gives a maximum yield. If you can find such a policy that has a larger annual yield than the uniform policy, then recommend it. On the other hand, shepherds who raise sheep for their wool may prefer to minimize the annual yield. If you can find a sustainable policy whose yield is smaller than that of the uniform policy, make a recommendation accordingly. In each case find the expected distribution of your harvesting policies. Do you think there are optimum harvesting policies of this type? Do you think that there might be other economic factors that should be taken into account in this model? Organize your results for a report to be read by your supervisor and an informed public.

Procedure: Express this problem as a discrete linear dynamical system $\mathbf{x}^{k+1} = L\mathbf{x}^k$, where L is a so-called Leslie matrix of the form

$$L = \begin{bmatrix} a_1 & a_2 & a_3 & \cdot & a_{n-1} & a_n \\ b_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & b_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & b_{n-1} & 0 \end{bmatrix}.$$

It is understood that the $0 < b_i \leq 1$, $0 \leq a_i$ and at least one $a_i \neq 0$. The facts you need to know (and may assume as standard facts about Leslie matrices) are as follows: such a matrix will have exactly one positive eigenvalue which turns out to be a simple eigenvalue (not repeated). Moreover, if at least two adjacent entries of the first row are positive, this eigenvalue will be a *dominant* eigenvalue, i.e., it is strictly larger than any other eigenvalue in absolute value. In particular, if the positive eigenvalue is 1, then starting from any nonzero initial state with nonnegative entries, successive states converge to an eigenvector belonging to the eigenvalue 1 which has all nonnegative entries. Scale this vector by dividing it by the sum of its components and one obtains an eigenvector which is a probability distribution vector, i.e., its entries are nonnegative and sum to 1. The entries of this vector give the long term distribution of the population in the various age classes.

In regards to harvesting, let H be a diagonal matrix with the harvest fractions h_i down the diagonal. (Here $0 \leq h_i \leq 1$.) Then the population that results from this harvesting at the end of each period is given by $\mathbf{x}^{k+1} = L\mathbf{x}^k - HL\mathbf{x}^k = (I - H)L\mathbf{x}^k$. But the matrix $(I - H)L$ is itself a Leslie matrix, so the theory applies to it as well. There are other theoretical tools, but all you need to do is to find a matrix $H = hI$ so that 1 is the positive eigenvalue of $(I - H)L$. You can do this by trial and error, a method which is applicable to any harvesting policy, uniform or not. However, in the case of uniform policies it's simpler to note that $(I - H)L = (1 - h)L$, where h is the diagonal entry of H.

Implementation Notes: (To the instructor: Add local notes here and discuss available aids. For example, when I give this assignment under Maple or Mathematica, I create a notebook that has the correct vector of a_i's and b_i's in it to avoid a very common problem: data entry error.)

5.7 Exercises

1. Let $D = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$ and $M = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. Compute $P = M^T M$ and $A = P^{-1}DP$. Explain why the eigenvalues of A are $-2, 1, 5$. Also explain why, given the form of M, we know that P^{-1} is sure to have integer entries before we even calculate P^{-1}.

2. A square matrix is said to be *strictly diagonally dominant* if in each row the sum of the absolute values of the off-diagonal entries is strictly less than the absolute value of the diagonal entry. Show that a strictly diagonally dominant matrix is invertible. *Hint:* Use Gershgorin to show that 0 is not an eigenvalue of the matrix.

3. The matrix A below may have complex eigenvalues.

$$A = \begin{bmatrix} 1 & -2 & -2 & 0 \\ 6 & -7 & 21 & -18 \\ 4 & -8 & 22 & -18 \\ 2 & -4 & 13 & -13 \end{bmatrix}$$

Use the Gerschgorin circle theorem to locate eigenvalues and the iteration methods of this section to compute an approximate eigensystem.

Review

Chapter 5 Exercises

1. Find the characteristic polynomial and eigenvalues of matrix $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & 4 \end{bmatrix}$.

2. For the matrix $A = \begin{bmatrix} 1 & 3 & 3 \\ 0 & 5 & 4 \\ 0 & 0 & 1 \end{bmatrix}$ find a matrix P and a diagonal matrix D such that $P^{-1}AP = D$. (You do *not* have to find P^{-1}).

3. Given that a 5×5 has only one eigenvalue λ, what are the possible Jordan canonical forms of the matrix?

4. Answer True/False:

(a) The matrix $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ is diagonalizable.

(b) The matrix $\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$ is diagonalizable.

(c) Every eigenvalue of the matrix A is nonzero.

(d) Every real matrix is similar to a diagonal matrix.

(e) If A is diagonalizable, then A^T is diagonalizable.

5. Verify directly from definition that if λ is an eigenvalue of A, then $\lambda + 1$ is an eigenvalue of $A + I$.

6. Find two matrices A, B such that the only eigenvalues of A and B are 0 but AB has nonzero eigenvalues.

7. Let $A = \mathbf{a}\mathbf{b}^T$ where $\mathbf{a}$ and $\mathbf{b}$ are $n \times 1$ column vectors. Find all eigenvalues and eigenvectors for A.

8. Show that if 0 is an eigenvalue of A, then A is not invertible

9. A matrix A is called *normal* if $A^H A = AA^H$. Prove that every Hermitian symmetric matrix is normal.

10. Let $A = \begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix}$. One of the eigenvalues of A is 0.

(a) Use the trace (Exercise 4) to find the other eigenvalue.

(b) Find eigenvectors for each eigenvalue.

(c) Unitarily orthogonalize the matrix A.

11. Show that if the matrix A is diagonalizable and has only one eigenvalue (repeated, of course), then A is a scalar matrix.

12. Show that if the matrix A is unitarily diagonalizable, then so is A^H. Prove or disprove that the same is true for A^T

CHAPTER 6

GEOMETRICAL ASPECTS OF ABSTRACT SPACES

Two basic ideas that we learn in geometry are that of length of a line segment and angle between lines. We have already seen how to extend the ideas to the standard vector spaces. The objective of this chapter is to extend these powerful ideas to general linear spaces. A surprising number of concepts and techniques that we learned in a standard setting can be carried over, almost word for word, to more general vector spaces. Once this is accomplished, we will be able to use our geometrical intuition in entirely new ways. For example, we will be able to have notions of length and perpendicularity for nonstandard vectors such as functions in a function space. Another application is that we will be able to give a sensible meaning to the size of the error incurred in solving a linear system with finite precision arithmetic. There are many more uses for this abstraction, as we shall see.

6.1. Normed Linear Spaces

Definitions and Examples

The basic function of a norm is to measure length and distance, independent of any other considerations, such as angles or orthogonality. There are different ways to accomplish such a measurement. One method of measuring length might be more natural for a given problem, or easier to calculate than another. For these reasons, we would like to have the option of using different methods of length measurement. You may recognize the properties listed below from earlier in the text; they are the basic norm laws given in Section 4.1 of Chapter 4 for the standard norm. We are going to abstract the norm idea to arbitrary vector spaces.

Norm Properties

DEFINITION 6.1.1. A *norm* on the vector space V is a function $||\cdot||$ which assigns to each vector $\mathbf{v} \in V$ a real number $||v||$ such that for c a scalar and $\mathbf{u}, \mathbf{v} \in V$ the following hold:

1. $||\,\mathbf{u}\,|| \geq 0$ with equality if and only if $\mathbf{u} = 0$.
2. $||\,c\mathbf{u}\,|| = |\,c\,|\; ||\,\mathbf{u}\,||$
3. (Triangle Inequality) $||\,\mathbf{u} + \mathbf{v}\,|| \leq ||\,\mathbf{u}\,|| + ||\,\mathbf{v}\,||$

DEFINITION 6.1.2. A vector space V, together with a norm $||\cdot||$ on the space V, is called a *normed linear space*.

Notice that if V is a normed linear space and W is any subspace of V, then W automatically becomes a normed linear space if we simply use the norm of V on elements of W. For obviously all the norm laws still hold, since they hold for elements of the bigger space V.

Of course, we have already studied some very important examples of normed linear spaces, namely the standard vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$, or any subspace thereof, together with the standard norms given by

$$\begin{aligned} ||(z_1, z_2, \ldots, z_n)|| &= \sqrt{z_1\overline{z_1} + z_2\overline{z_2} + \cdots + z_n\overline{z_n}} \\ &= \left(|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2\right)^{1/2} \end{aligned}$$

If the vectors are real then we can drop the conjugate bars. This norm is actually one of a family of norms which are commonly used.

DEFINITION 6.1.3. Let V be one of the standard spaces $\mathbb{R}^n$ or $\mathbb{C}^n$ and $p \geq 1$ a real number. The p-norm of a vector in V is defined by the formula p-Norms

$$||(z_1, z_2, \ldots, z_n)||_p = (|z_1|^p + |z_2|^p + \cdots + |z_n|^p)^{1/p}$$

Notice that when $p = 2$ we have the familiar example of the standard norm. Another important case is that in which $p = 1$ which gives (not surprisingly) the so-called 1-norm. The last important instance of a p-norm is one that isn't so obvious: $p = \infty$. It turns out that the value of this norm is the limit of p-norms as $p \to \infty$. To keep matters simple, we'll supply a separate definition for this norm.

DEFINITION 6.1.4. Let V be one of the standard spaces $\mathbb{R}^n$ or $\mathbb{C}^n$. The ∞-norm of a vector in V is defined by the formula

$$||(z_1, z_2, \ldots, z_n)||_\infty = \max\{|z_1|, |z_2|, \ldots, |z_n|\}$$

EXAMPLE 6.1.5. Calculate $||\mathbf{v}||_p$ where $p = 1, 2$ or ∞ and $\mathbf{v} = (1, -3, 2, -1) \in \mathbb{R}^4$.

SOLUTION. We calculate:

$$\begin{aligned} ||(1,-3,2,-1)||_1 &= |1| + |-3| + |2| + |-1| = 7 \\ ||(1,-3,2,-1)||_2 &= \sqrt{|1|^2 + |-3|^2 + |2|^2 + |-1|^2} = \sqrt{15} \\ ||(1,-3,2,-1)||_\infty &= \max\{|1|, |-3|, |2|, |-1|\} = 3 \end{aligned}$$

□

It may seem a bit odd at first to speak of the same vector as having different lengths. You should take the point of view that choosing a norm is a bit like choosing a measuring stick. If you choose a yard stick, you won't measure the same number as you would by using a meter stick on the same object.

EXAMPLE 6.1.6. Verify that the norm properties are satisfied for the p-norm in the case that $p = \infty$.

SOLUTION. Let c be a scalar, $\mathbf{u} = (z_1, z_2, \ldots, z_n)$ and $\mathbf{v} = (w_1, w_2, \ldots, w_n)$ two vectors. Any absolute value is nonnegative and any vector whose largest component in absolute value is zero must have all components equal to zero. Property (1) follows. Next, we have that

$$\begin{aligned} ||c\mathbf{u}||_\infty &= ||(cz_1, cz_2, \ldots, cz_n)||_\infty \\ &= \max\{|cz_1|, |cz_2|, \ldots, |cz_n|\} \\ &= |c| \max\{|z_1|, |z_2|, \ldots, |z_n|\} = |c|\ ||\mathbf{u}||_\infty \end{aligned}$$

which proves (2). For (3) we observe that

$$\begin{aligned} ||\mathbf{u}+\mathbf{v}||_\infty &= \max\{|z_1| + |w_1|, |z_2| + |w_2|, \ldots, |z_n| + |w_n|\} \\ &\leq \max\{|z_1|, |z_2|, \ldots, |z_n|\} + \max\{|w_1|, |w_2|, \ldots, |w_n|\} \\ &\leq ||\mathbf{u}||_\infty + ||\mathbf{v}||_\infty \end{aligned}$$

□

Unit Vectors

Sometimes it is convenient to deal with vectors whose length is one. Such a vector is called a *unit vector*. We saw in Chapter 3 that it is easy to concoct a unit vector $\mathbf{u}$ in the same direction as a given nonzero vector $\mathbf{v}$ when using the standard norm, namely take

$$\mathbf{u} = \frac{\mathbf{v}}{||\mathbf{v}||} \tag{6.1.1}$$

The same formula holds for any norm whatsoever because of norm property (2).

EXAMPLE 6.1.7. Construct a unit vector in the direction of $\mathbf{v} = (1, -3, 2, -1)$, where the 1-norm, 2-norm, and ∞-norms are used to measure length.

SOLUTION. We already calculated each of the norms of $\mathbf{v}$ in Example 6.1.5. Use these numbers in Equation 6.1.1 to obtain unit length vectors

$$\begin{aligned} \mathbf{u}_1 &= \frac{1}{7}(1, -3, 2, -1) \\ \mathbf{u}_2 &= \frac{1}{\sqrt{15}}(1, -3, 2, -1) \\ \mathbf{u}_\infty &= \frac{1}{3}(1, -3, 2, -1) \end{aligned}$$

□

From a geometric point of view there are certain sets of vectors in the vector space V that tell us a lot about distances. These are the so-called *balls about a vector (or point)* $\mathbf{v}_0$ *of radius* r whose definition is as follows:

$$B_r(\mathbf{v}_0) = \{\mathbf{v} \in V \mid ||\mathbf{v} - \mathbf{v}_0|| \leq r\}$$

Sometimes these are called *closed balls*, as opposed to *open balls* which are defined by using strict inequality. Here is a situation in which these balls are very helpful: imagine trying to find the distance from a given vector $\mathbf{v}_0$ to a closed (this means it contains all points on its boundary) set S of vectors which need not be a subspace. One way to accomplish this is to start with a ball centered at $\mathbf{v}_0$ such that the ball avoids S. Then expand this ball by increasing its radius until you have found a least radius r such that the ball $B_r(\mathbf{v}_0)$ intersects S nontrivially. Then the distance from $\mathbf{v}_0$ to this set is this number r. Actually, this is a reasonable *definition* of the distance from $\mathbf{v}_0$ to the set S. One expects these balls, for a given norm, to have the same shape, so it is sufficient to look at the unit balls, that is, the case where $r = 1$.

EXAMPLE 6.1.8. Sketch the unit balls centered at the origin for the 1-norm, 2-norm, and ∞-norms in the space $V = \mathbb{R}^2$.

SOLUTION. In each case it's easiest to determine the boundary of the ball $B_1(0)$, i.e., the set of vectors $\mathbf{v} = (x, y)$ such that $||\mathbf{v}|| = 1$. These boundaries are sketched in Figure 6.1.1 and the ball consists of the boundaries plus the interior of each boundary. Let's start with the familiar norm 2-norm. Here the boundary consists of points (x, y) such that

$$1 = ||(x, y)||_2 = x^2 + y^2$$

which is the familiar circle of radius 1 centered at the origin. Next, consider the 1-norm in which case

$$1 = ||(x, y)||_1 = |x| + |y|$$

It's easier to examine this formula in each quadrant, where it becomes one of the four possibilities

$$\pm x \pm y = 1$$

For example, in the first quadrant we get $x + y = 1$. These equations give lines which connect to from a square whose sides are diagonal lines. Finally, for the ∞-norm we have

$$1 = ||(x, y)||_\infty = \max\{|x|, |y|\}$$

which gives four horizontal and vertical lines $x = \pm 1$ and $y = \pm 1$ which intersect to form another square. Thus we see that the unit "balls" for the 1- and ∞-norms have corners, unlike the 2-norm. see Figure 6.1.1 for a picture of these balls. □

One more comment about norms. Recall from Section 4.1 that one of the important applications of the norm concept is that it enables us to make sense out of the idea of limits of vectors. In a nutshell $\lim_{n\to\infty} \mathbf{v}_n = \mathbf{v}$ was taken to mean that $\lim_{n\to\infty} ||\mathbf{v}_n - \mathbf{v}|| = 0$. Will we have to have a different notion of limits for different norms? The somewhat surprising answer is "no." The reason is that given any two norms $||\cdot||_a$ and $||\cdot||_b$ on a finite dimensional vector space, it is always possible to find positive real constants c and d such that for any vector $\mathbf{v}$ Limit of Vectors

$$||\mathbf{v}||_a \le c \cdot ||\mathbf{v}||_b \text{ and } ||\mathbf{v}||_b \le d \cdot ||\mathbf{v}||_a$$

Hence, if $||\mathbf{v}_n - \mathbf{v}||$ tends to 0 in one norm, it will tend to 0 in the other norm. In this sense, it can be shown that all norms on a finite dimensional vector space are equivalent. Indeed, it can be shown that the condition that $||\mathbf{v}_n - \mathbf{v}||$ tends to 0 in any one norm is equivalent to the condition that each coordinate of $\mathbf{v}_n$ converge to the corresponding coordinate of $\mathbf{v}$. We will verify the limit fact in the following example.

EXAMPLE 6.1.9. Verify that $\lim_{n\to\infty} \mathbf{v}_n$ exists and is the same with respect to both the 1-norm and 2-norm, where

$$\mathbf{v}_n = \begin{bmatrix} (1-n)/n \\ e^{-n} + 1 \end{bmatrix}$$

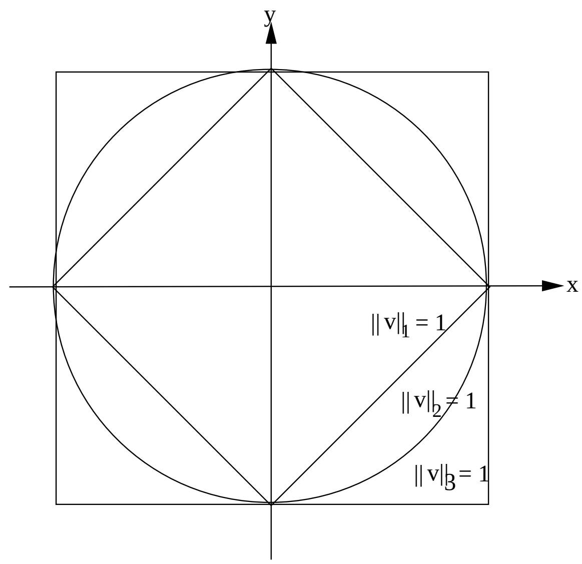


FIGURE 6.1.1. Boundaries of unit balls in various norms.

Which norm is easier to work with?

SOLUTION. First we have to know what the limit will be. Let's examine the limit in each coordinate. We have

$$\lim_{n\to\infty}\frac{1-n}{n}=\lim_{n\to\infty}\frac{1}{n}-1=0-1=-1 \quad \text{and} \quad \lim_{n\to\infty}e^{-n}+1=0+1=1.$$

So we will try to use $\mathbf{v}=(-1,1)$ as the limiting vector. Now calculate

$$\mathbf{v}-\mathbf{v}_n=\begin{bmatrix}-1\\1\end{bmatrix}-\begin{bmatrix}\frac{1-n}{n}\\e^{-n}+1\end{bmatrix}=\begin{bmatrix}\frac{1}{n}\\e^{-n}\end{bmatrix}$$

so that

$$||\mathbf{v}-\mathbf{v}_n||_1=|\frac{1}{n}|+|e^{-n}|\underset{n\to\infty}{\longrightarrow}0,$$

and

$$||\mathbf{v}-\mathbf{v}_n||=\sqrt{\left(\frac{1}{n}\right)^2+(e^{-n})^2}\underset{n\to\infty}{\longrightarrow}0$$

which shows that the limits are the same in either norm. In this case the 1-norm appears to be easier to work with since no squaring and square roots are involved. □

6.1 Exercises

1. Let $\mathbf{x}=[1+i,-1,0,1]^T$ and $\mathbf{y}=[1,1,2,-4]^T$ be vectors in $\mathbb{C}^4$. Find the 1- 2- and ∞- norms of the vectors $\mathbf{x}$ and $\mathbf{y}$.

2. Find unit vectors in the direction of $\mathbf{v}=(1,-3,-1)$ with respect to the 1-, 2-, and ∞-norms.

3. Verify the triangle inequality for $\mathbf{u} = (0, 2, 3, 1)$, $\mathbf{v} = (1, -3, 2, -1)$ and the 1-norm.

4. Verify that all three of the norm laws hold in the case that $c = -3$, $\mathbf{u} = (2, -4, 1)$, $\mathbf{v} = (1, 2, -1)$ and the norm is the infinity norm.

5. Find the distance from the origin to the line $x+y = 2$ using the ∞-norm by sketching a picture of the ball centered at the origin that touches the line.

6. Given the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, find the largest possible value of $||A\mathbf{x}||_\infty$, where $\mathbf{x}$ ranges over the vectors whose ∞-norm is 1.

7. Verify that the 1-norm satisfies the definition of a norm.

8. Verify that $\lim_{n\to\infty} \mathbf{v}_n$ exists and is the same with respect to both the 1- and 2-norm, where

$$\mathbf{v}_n = \begin{bmatrix} (1-n)/n \\ e^{-n} + 1 \end{bmatrix}$$

9. Calculate $\lim_{n\to\infty} \mathbf{v}_n$ using the ∞-norm, where

$$\mathbf{v}_n = \begin{bmatrix} n^2/(n^3+1) \\ \sin(n)/(n^3+1) \end{bmatrix}$$

10. An example of a norm on $R^{m,n}$ (or $\mathbb{C}^{m,n}$) is the *Frobenius norm* of an $m \times n$ matrix $A = [a_{ij}]$ is defined by the formula

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}.$$

Compute the Frobenius norm of the following matrices.

(a) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 2 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & -2 \end{bmatrix}$ (c) $\begin{bmatrix} 1+2i & 2 \\ 1 & -3i \end{bmatrix}$

11. Show that the Frobenius norm satisfies norm properties 1 and 2.

6.2. Inner Product Spaces

Definitions and Examples

We saw in Section 4.2 that the notion of a dot product of two vectors had many handy applications, including the determination of the angle between two vectors. This dot product amounted to the “concrete” inner product of the two standard vectors. We now extend this idea in a setting that allows for real or complex abstract vector spaces.

DEFINITION 6.2.1. An (abstract) *inner product* on the vector space V is a function $\langle \cdot, \cdot \rangle$ which assigns to each pair of vectors $\mathbf{u}, \mathbf{v} \in V$ a scalar $\langle \mathbf{u}, \mathbf{v} \rangle$ such that for c a scalar and $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ the following hold:

1. $\langle \mathbf{u}, \, \mathbf{u} \rangle \geq 0$ with equality if and only if $\mathbf{u} = 0$
2. $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$
3. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
4. $\langle \mathbf{u}, c\mathbf{v} \rangle = c \, \langle \mathbf{u}, \mathbf{v} \rangle$

DEFINITION 6.2.2. A vector space V, together with an inner product on the space V, is called an *inner product space*.

Notice that in the case of the more common vector spaces over *real scalars*, property 2 becomes a commutativity law: $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.

Also observe that if V is an inner product space and W is any subspace of V, then W automatically becomes an inner product space if we simply use the inner product of V on elements of W. For obviously all the inner product laws still hold, since they hold for elements of the bigger space V.

Of course, we have the standard examples of inner products, namely the dot products on $\mathbb{R}^n$ and $\mathbb{C}^n$. Here is an example of a nonstandard inner product that is useful in certain engineering problems.

EXAMPLE 6.2.3. For vectors $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ in $V = \mathbb{R}^2$, define an inner product by the formula

$$\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + 3u_2v_2$$

Show that this formula satisfies the definition of inner product.

SOLUTION. First we see that

$$\langle \mathbf{u}, \mathbf{u} \rangle = 2u_1^2 + 3u_2^2$$

so the only way for this sum to be 0 is for $u_1 = u_2 = 0$. Hence (1) holds. For (2) calculate

$$\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + 3u_2v_2 = 2v_1u_1 + 3v_2u_2 = \langle \mathbf{v}, \mathbf{u} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$$

since all scalars in question are real. For (3) let $\mathbf{w} = (w_1, w_2)$ and calculate

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle &= 2u_1(v_1 + w_1) + 3u_2(v_2 + w_2) \\ &= 2u_1v_1 + 3u_2v_2 + 2u_1w_1 + 3u_2 \\ &= \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle \end{aligned}$$

For the last property, check that for a scalar c

$$\begin{aligned} \langle \mathbf{u}, c\mathbf{v} \rangle &= 2u_1cv_1 + 3u_2cv_2 \\ &= c(2u_1v_1 + 3u_2v_2) \\ &= c \, \langle \mathbf{u}, \mathbf{v} \rangle \end{aligned}$$

$\square$

It follows that this "weighted" inner product is indeed an inner product according to our definition. In fact, we can do a whole lot more with even less effort. Consider this example, of which the preceding is a special case.

EXAMPLE 6.2.4. Let $V = \mathbb{R}^n$ or $\mathbb{C}^n$ and let $\mathbf{u}, \mathbf{v} \in V$. Let A be a fixed $n \times n$ *nonsingular* matrix. Show that the matrix A defines an inner product by the formula

$$\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u})^H (A\mathbf{v}) = \mathbf{u}^H A^H A\mathbf{v}.$$

SOLUTION. As usual, let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and let c be a scalar. Use the norm sign for the ordinary 2-norm and we have

$$\langle \mathbf{u}, \mathbf{u} \rangle = ||A\mathbf{u}||^2$$

so that if $\langle \mathbf{u}, \mathbf{u} \rangle = 0$, then $A\mathbf{u} = 0$. We are assuming A is nonsingular, so that this implies that $\mathbf{u} = 0$, which establishes property (1). For (2), remember that for a 1×1 scalar quantity q, $q^H = \overline{q}$ so we calculate:

$$\langle \mathbf{v}, \mathbf{u} \rangle = \mathbf{v}^H A^H A\mathbf{u} = (\mathbf{u}^H A^H A\mathbf{v})^H = \langle \mathbf{u}, \mathbf{v} \rangle^H = \overline{\langle \mathbf{u}, \mathbf{v} \rangle}$$

Next, we have

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle &= \mathbf{u}^H A^H A(\mathbf{v} + \mathbf{w}) \\ &= \mathbf{u}^H A^H A\mathbf{v} + \mathbf{u}^H A^H A\mathbf{w} \\ &= \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle \end{aligned}$$

Finally, we have that

$$\begin{aligned} \langle \mathbf{u}, c\mathbf{v} \rangle &= \mathbf{u}^H A^H Ac\mathbf{v} \\ &= c\mathbf{u}^H A^H A\mathbf{v} \\ &= c \langle \mathbf{u}, \mathbf{v} \rangle \end{aligned}$$

This shows that the inner product properties are satisfied. □

We leave it to the reader to check that if we take

$$A = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{3} \end{bmatrix}$$

then the inner product defined by this matrix is exactly the inner product of Example 6.2.3.

There is an important point to be gleaned from the previous example, namely, that a given vector space may have more than one inner product on it. In particular, $V = \mathbb{R}^2$ could have the standard inner product – dot products – or something like the previous example. The space V, together with each one of these inner products, provide us with two separate inner product spaces.

Here is a rather more exotic example of an inner product, in that it does not involve one of standard spaces for its underlying vector space.

EXAMPLE 6.2.5. Let $V = C[0, 1]$, the space of continuous functions on the interval $[0, 1]$ with the usual function addition and scalar multiplication. Show that the formula

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx$$

defines an inner product on the space V.

SOLUTION. Certainly $\langle f, g\rangle$ is a real number. Now if $f(x)$ is a continuous function then $f(x)^2$ is nonnegative on $[0,1]$ and therefore $\int_0^1 f(x)^2 dx = \langle f, f\rangle \geq 0$. Furthermore, if $f(x)$ is nonzero, then the area under the curve $y = f(x)^2$ must also be positive since $f(x)$ will be positive and bounded away from 0 on some subinterval of $[0,1]$. This establishes property (1) of inner products.

Now let $f(x), g(x), h(x) \in V$. For property (2), notice that

$$\langle f, g\rangle = \int_0^1 f(x)g(x)dx = \int_0^1 g(x)f(x)dx = \langle g, f\rangle .$$

Also,

$$\begin{aligned}\langle f, g+h\rangle &= \int_0^1 f(x)(g(x)+h(x))dx \\ &= \int_0^1 f(x)g(x)dx + \int_0^1 f(x)h(x)dx \\ &= \langle f, g\rangle + \langle f, h\rangle\end{aligned}$$

which establishes property (3). Finally, we see that for a scalar c

$$\langle f, cg\rangle = \int_0^1 f(x)cg(x)\,dx = c\int_0^1 f(x)g(x)\,dx = c\,\langle f, g\rangle$$

which shows that property (4) holds. $\square$

Clearly we could similarly define an inner product on the space $C[a,b]$ of continuous functions on any finite interval $[a,b]$ just as in the preceding example by changing the limits of the integrals from $[0,1]$ to $[a,b]$. We shall refer to this inner product on a function space as the *standard inner product* on a function space.

Following are a few simple facts about inner products that we will use frequently. The proofs are left to the exercises.

THEOREM 6.2.6. *Let V be an inner product space with inner product $\langle\cdot,\cdot\rangle$. Then we have that for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and scalars a*

1. $\langle 0, \mathbf{u}\rangle = 0$
2. $\langle \mathbf{u}+\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle$
3. $\langle a\mathbf{u}, \mathbf{v}\rangle = \overline{a}\langle \mathbf{u}, \mathbf{v}\rangle$

Induced Norms and the CBS Inequality

It is a striking fact that we can accomplish all the goals we set for the standard inner product with general inner products as well: we can introduce the ideas of angles, orthogonality, projections and so forth. We have already seen much of the work that has to be done, though it was stated in the context of the standard inner products. As a first step, we want to point out that every inner product has a "natural" norm associated with it.

DEFINITION 6.2.7. Let V be an inner product space. For vectors $\mathbf{u} \in V$, the norm defined by the equation

$$||\mathbf{u}|| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$$

is called the *norm induced by the inner product* $\langle \cdot, \cdot \rangle$ *on* V.

Induced Norm

As a matter of fact, this idea is not really new. Recall that we introduced the standard inner product on $V = V = \mathbb{R}^n$ or $\mathbb{C}^n$ with an eye towards the standard norm. At the time it seemed like a nice convenience that the norm could be expressed in terms of the inner product. It is, and so much so that we have turned this cozy relationship into a definition. Is the induced norm really a norm? We have some work to do. The first norm property is easy to verify for the induced norm: from property (1) of inner products we see that $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, with equality if and only if $\mathbf{u} = 0$. This confirms norm property (1). Norm property (2) isn't too hard either: let c be a scalar and check that

$$||c\mathbf{u}|| = \sqrt{\langle c\mathbf{u}, c\mathbf{u} \rangle} = \sqrt{c\overline{c}\,\langle \mathbf{u}, \mathbf{u} \rangle} = \sqrt{|c|^2}\sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = |c|\;||\mathbf{u}||$$

Norm property (3), the triangle inequality, remains. This one isn't easy to verify from first principles. We need a tool that we have seen before, the Cauchy-Bunyakovsky-Schwarz (CBS) inequality. We restate it below as the next theorem. Indeed, the very same proof that is given in Theorem 4.2.1 of Chapter 3 carries over word for word to general inner products over real vector spaces. We need only replace dot products $\mathbf{u} \cdot \mathbf{v}$ by abstract inner products $\langle \mathbf{u}, \mathbf{v} \rangle$. Similarly, the proof of the triangle inequality as given in Example 4.2.4 of Chapter 3, carries over to establish the triangle inequality for abstract inner products. Hence property (3) of norms holds for any induced norm.

CBS Inequality

THEOREM 6.2.8. *Let V be an inner product space. For $\mathbf{u}, \mathbf{v} \in V$, if we use the inner product of V and its induced norm, then*

$$|\,\langle\, \mathbf{u}, \mathbf{v} \rangle\,| \leq ||\,\mathbf{u}\,||\;||\,\mathbf{v}\,||$$

Just as with the standard dot products, thanks to this inequality, we can formulate the following definition.

Angle Between Vectors

DEFINITION 6.2.9. For vectors $\mathbf{u}, \mathbf{v} \in V$, an inner product space, we define the *angle* between $\mathbf{u}$ and $\mathbf{v}$ to be any angle θ satisfying

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\,\mathbf{u}\,||\;||\,\mathbf{v}\,||}$$

We know that $|\,\langle \mathbf{u}, \mathbf{v} \rangle\,|/(||\,\mathbf{u}\,||\;||\,\mathbf{v}\,||) \leq 1$ so that this formula for $\cos\theta$ makes sense.

EXAMPLE 6.2.10. Let $\mathbf{u} = (1, -1)$ and $\mathbf{v} = (1, 1)$ be vectors in $\mathbb{R}^2$. Compute an angle between these two vectors using the inner product of Example 6.2.3. Compare this to angle found when one uses the standard inner product in $\mathbb{R}^2$.

SOLUTION. According to 6.2.3 and the definition of angle, we have

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\mathbf{u}|| \cdot ||\mathbf{v}||} = \frac{2 \cdot 1 \cdot 1 + 3 \cdot (-1) \cdot 1}{\sqrt{2 \cdot 1^2 + 3 \cdot (-1)^2}\sqrt{2 \cdot 1^2 + 3 \cdot 1^2}} = \frac{-1}{5}$$

Hence the angle in radians is

$$\theta = \arccos(\frac{-1}{5}) \approx 1.7722$$

On the other hand, if we use the standard norm then

$$\langle \mathbf{u}, \mathbf{v} \rangle = 1 \cdot 1 + (-1) \cdot 1 = 0$$

from which it follows that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal and $\theta = \pi/2 \approx 1.5708$. □

In the previous example, it shouldn't be too surprising that we can arrive at two different values for the "angle" between two vectors. Using different inner products to measure angle is somewhat like measuring length with different norms. Next, we extend the perpendicularity idea to arbitrary inner product spaces.

DEFINITION 6.2.11. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in the same inner product space are *orthogonal* if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Note that if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, then $\langle \mathbf{v}, \mathbf{u} \rangle = \overline{\langle \mathbf{u}, \mathbf{v} \rangle} = 0$. Also, this definition makes the zero vector orthogonal to every other vector. It also allows us to speak of things like "orthogonal functions." One has to be careful with new ideas like this. Orthogonality in a function space is not something that can be as easily visualized as orthogonality of geometrical vectors. Inspecting the graphs of two functions may not be quite enough. If, however, graphical data is tempered with a little understanding of the particular inner product in use, orthogonality can be detected.

EXAMPLE 6.2.12. Show that $f(x) = x$ and $g(x) = x - 2/3$ are orthogonal elements of $C[0, 1]$ with the inner product of Example 6.2.5 and provide graphical evidence of this fact.

SOLUTION. According to the definition of inner product in this space,

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx = \int_0^1 x(x - \frac{2}{3})dx = (\frac{x^3}{3} - \frac{x^2}{3})\Big|_0^1 = 0.$$

It follows that f and g are orthogonal to each other. For graphical evidence, sketch $f(x)$, $g(x)$ and $f(x)g(x)$ on the interval $[0, 1]$ as in Figure 6.2.1. The graphs of f and g are not especially enlightening; but we can see in the graph that the area below $f \cdot g$ and above the x-axis to the right of $(2/3, 0)$ seems to be about equal to the area to the left of $(2/3, 0)$ above $f \cdot g$ and below the x-axis. Therefore the integral of the product on the interval $[0, 1]$ might be expected to be zero, which is indeed the case. □

Some of the basic ideas from geometry that fuel our visual intuition extend very elegantly to the inner product space setting. One such example is the famous Pythagorean Theorem, which takes the following form in an inner product space.

Pythagorean Theorem

THEOREM 6.2.13. *Let* $\mathbf{u}, \mathbf{v}$ *be orthogonal vectors in an inner product space* V. *Then* $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u} + \mathbf{v}\|^2$.

PROOF. Compute

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \end{aligned}$$

□

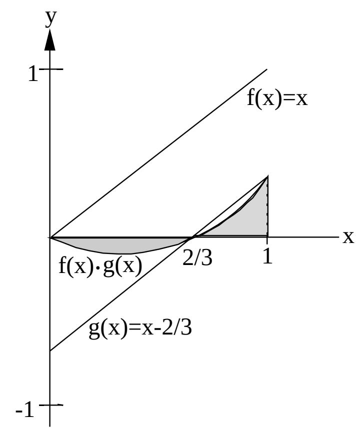


FIGURE 6.2.1. Graph of f,g and $f \cdot g$ on the interval $[0, 1]$.

Here is an example of another standard geometrical fact that fits well in the abstract setting. This is equivalent to the law of parallelograms, which says that the sum of the squares of the diagonals of a parallelogram is equal to the sum of the squares of all four sides.

EXAMPLE 6.2.14. Use properties of inner products to show that if we use the induced norm, then

$$||\mathbf{u}+\mathbf{v}||^2 + ||\mathbf{u}-\mathbf{v}||^2 = 2\left(||\mathbf{u}||^2 + ||\mathbf{v}||^2\right)$$

SOLUTION. The key to proving this fact is to relate induced norm to inner product. Specifically,

$$\|\mathbf{u}+\mathbf{v}\|^2 = \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle$$

while

$$\|\mathbf{u}-\mathbf{v}\|^2 = \langle \mathbf{u}-\mathbf{v}, \mathbf{u}-\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{u}\rangle - \langle \mathbf{u}, \mathbf{v}\rangle - \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle$$

Now add these two equations and obtain by using the definition of induced norm again that

$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\langle \mathbf{u}, \mathbf{u}\rangle + 2\langle \mathbf{v}, \mathbf{v}\rangle = 2(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2)$$

which is what was to be shown. □

It would be nice to think that every norm on a vector space is induced from some inner product. Unfortunately, this is not true, as the following example shows.

EXAMPLE 6.2.15. Use the result of Example 6.2.14 to show that the infinity norm on $V = \mathbb{R}^2$ is not induced by any inner product on V.

SOLUTION. Suppose the infinity norm were induced by some inner product on V. Let $\mathbf{u} = (1, 0)$ and $\mathbf{v} = (0, 1/2)$. Then we have

$$||\mathbf{u}+\mathbf{v}||_\infty^2 + ||\mathbf{u}-\mathbf{v}||_\infty^2 = ||(1, 1/2)||_\infty^2 + ||(1, -1/2)||_\infty^2 = 2$$

while

$$2\left(||\mathbf{u}||^2 + ||\mathbf{v}||^2\right) = 2(1 + 1/4) = 5/2$$

This contradicts Example 6.2.14, so that the infinity norm cannot be induced from an inner product. □

One last example of a geometrical idea that generalizes to inner product spaces is the notion of projections of one vector along another. The projection formula for vectors of Chapter 4 works perfectly well for general inner products. Since the proof of this fact amounts to replacing dot products by inner products in the original formulation of the theorem (see page 273), we omit it and simply state the result.

Projection Formula for Vectors

THEOREM 6.2.16. *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be vectors in an inner product space with* $\mathbf{v} \neq 0$. *Define the projection of* $\mathbf{u}$ *along* $\mathbf{v}$ *as*

$$\mathbf{p} = \operatorname{proj}_{\mathbf{v}} \mathbf{u} \equiv \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}$$

and let $\mathbf{q} = \mathbf{u} - \mathbf{p}$. *Then* $\mathbf{p}$ *is parallel to* $\mathbf{v}$, $\mathbf{q}$ *is orthogonal to* $\mathbf{v}$ *and* $\mathbf{u} = \mathbf{p} + \mathbf{q}$.

Orthogonal Sets of Vectors

We have already seen the development of the ideas of orthogonal sets of vectors and bases in Chapter 4. Much of this development can be abstracted easily to general inner product spaces, simply by replacing dot products by inner products. Accordingly, we can make the following definition.

DEFINITION 6.2.17. The set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in an inner product space are said to be an *orthogonal set* if $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ whenever $i \neq j$. If, in addition, each vector has unit length, i.e., $\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1$ then the set of vectors is said to be an *orthonormal set* of vectors.

The proof of the following key fact and its corollary are the same as that of Theorem 6.2.18 in Section 4.3 of Chapter 4. All we have to do is replace dot products by inner products. The observations that followed the proof of this theorem are valid for general inner products as well. We omit the proofs and refer the reader to Chapter 4.

Orthogonal Coordinates Theorem

THEOREM 6.2.18. *Let* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *be an orthogonal set of nonzero vectors and suppose that* $\mathbf{v} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. *Then* $\mathbf{v}$ *can be expressed uniquely (up to order) as a linear combination of* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, *namely*

$$\mathbf{v} = \frac{\langle \mathbf{v}_1, \mathbf{v} \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{v}_2, \mathbf{v} \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 + \ldots + \frac{\langle \mathbf{v}_n, \mathbf{v} \rangle}{\langle \mathbf{v}_n, \mathbf{v}_n \rangle}\mathbf{v}_n$$

COROLLARY 6.2.19. *Every orthogonal set of nonzero vectors is linearly independent.*

Another useful corollary is the following fact about length of a vector whose proof is left as an exercise.

COROLLARY 6.2.20. *If* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *is an orthonormal set of vectors and*

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n$$

then

$$||\mathbf{v}||^2 = c_1^2 + c_2^2 + \ldots + c_n^2$$

EXAMPLE 6.2.21. Turn the orthogonal set $\{(1,-1,0),(1,1,0),(0,0,1)\}$ into an orthonormal set, calculate the coordinates of the vector $\mathbf{v} = (2,-1,-1)$ with respect to this orthonormal set and verify the formula just given for the length of $\mathbf{v}$.

SOLUTION. From the example we have $\mathbf{v}_1 = (1,-1,0), \mathbf{v}_2 = (1,1,0)$, and $\mathbf{v}_3 = (0,0,1)$. We see that $\langle \mathbf{v}_1, \mathbf{v}_1 \rangle = 2 = \langle \mathbf{v}_2, \mathbf{v}_2 \rangle$ and $\langle \mathbf{v}_3, \mathbf{v}_3 \rangle = 1$. So set $\mathbf{u}_1 = (1/\sqrt{2})\mathbf{v}_1$, $\mathbf{u}_2 = (1/\sqrt{2})\mathbf{v}_2$, and $\mathbf{u}_3 = \mathbf{v}_3$ to obtain an orthonormal set of vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$. Now the coordinates of $\mathbf{v}$ are easily calculated:

$$\begin{aligned}
c_1 &= \langle \mathbf{u}_1, \mathbf{v} \rangle = \frac{1}{\sqrt{2}}(1 \cdot 2 + (-1) \cdot (-1) + 0 \cdot (-1)) = \frac{3}{\sqrt{2}} \\
c_2 &= \langle \mathbf{u}_2, \mathbf{v} \rangle = \frac{1}{\sqrt{2}}(1 \cdot 2 + 1 \cdot (-1) + 0 \cdot (-1)) = \frac{1}{\sqrt{2}} \\
c_3 &= \langle \mathbf{u}_3, \mathbf{v} \rangle = 0 \cdot 2 + 0 \cdot (-1) + 1 \cdot (-1) = -1
\end{aligned}$$

from which we conclude that

$$\mathbf{v} = \frac{3}{\sqrt{2}}\mathbf{u}_1 + \frac{1}{\sqrt{2}}\mathbf{u}_2 - \mathbf{u}_3 = \frac{3}{2}\mathbf{v}_1 + \frac{1}{2}\mathbf{v}_2 - \mathbf{v}_3$$

Now from definition we have that $||\mathbf{v}||^2 = 2^2 + (-1)^2 + (-1)^2 = 6$ while $c_1^2 + c_2^2 + c_3^2 = 9/2 + 1/2 + 1 = 6$ as well. This confirms that the length squared of $\mathbf{v}$ is the sum of squares of the coordinates of $\mathbf{v}$ with respect to an orthonormal basis. $\square$

6.2 Exercises

1. Verify the Cauchy-Bunyakovsky-Schwarz inequality for $\mathbf{u} = (1,2)$ and $\mathbf{v} = (1,-1)$ using the weighted inner product on $\mathbb{R}^2$ given by $<(x,y),(w,z)> = 2xw + 3yz$.

2. Find the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$ in the following:

(a) $\mathbf{u}$ and $\mathbf{v}$ in Exercise 1 with the inner product given therein.

(b) $\mathbf{u} = x$ and $\mathbf{v} = x^3$ in $V = C[0,1]$ with the standard inner product as in Example 6.2.5.

3. Which of the following sets of vectors are linearly independent? Orthogonal? Orthonormal?

(a) $(1,-1,2),(2,2,0)$ in $\mathbb{R}^2$ with the standard inner product.

(b) $1, x, x^2$ as vectors in $C[-1,1]$ with the standard inner product on the interval $[-1,1]$.

(c) $\frac{1}{5}(3,4), \frac{1}{5}(4,-3)$ in $\mathbb{R}^2$ with the standard inner product.

(d) 1,$\cos(x)$, $\sin(x)$ in $C[-\pi,\pi]$ with the standard inner product.

(e) $(2,4),(1,0)$ in $\mathbb{R}^2$ with inner product (assume it is) $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \mathbf{y}$.

4. The law $<[x_1,x_2]^T,[y_1,y_2]^T>= 3x_1y_1 - 2x_2y_2$ fails to define an inner product on $\mathbb{R}^2$. Why?

5. If the square real matrix A has a nonpositive real eigenvalue, then the formula $\langle \mathbf{u},\mathbf{v}\rangle = \mathbf{u}^T A\mathbf{v}$ does not define an inner product. Why? *Hint:* Start with the definition of eigenvalue.

6. Show that the law $<(x_1,x_2),(y_1,y_2)>= x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2$ defines an inner product on $\mathbb{R}^2$. (It helps to know that $\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$.)

7. Verify that the inner product of Example 6.2.3 can be defined by using the matrix

$$A = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{3} \end{bmatrix}$$

together with the definition of matrix defined inner products from Example 6.2.4.

8. If $A = \begin{bmatrix} 3 & 1 \\ -1 & 2 \end{bmatrix}$, find the cosine of the angle θ between the two vectors $\mathbf{u} = [1,0]^T$ and $\mathbf{v} = [0,1]^T$ in the vector space $\mathbb{R}^2$ with respect to the inner product defined by

$$\langle \mathbf{u},\mathbf{v}\rangle \equiv (A\mathbf{u})^T(A\mathbf{v}) = \mathbf{u}^T(A^TA)\mathbf{v}.$$

9. Explain how one could use Theorem 4.3.3 to test for whether or not a given vector $\mathbf{w}$ in the inner product space W belongs to a subspace V and illustrate it by determining if $\mathbf{w} = (2,-4,3)$ belongs to the subspace of $\mathbb{R}^3$ spanned by $\mathbf{v}_1 = (1,1,0)$ and $\mathbf{v}_2 = (-1,1,1)$.

10. Prove that $||\cdot||_1$ is not an induced norm on $\mathbb{R}^n$. *Hint:* See Example 6.2.15.

11. Let A be an $n \times n$ real matrix and define the product $\langle \mathbf{u},\mathbf{v}\rangle \equiv \mathbf{u}^T A\mathbf{v}$ for all $\mathbf{u},\mathbf{v} \in \mathbb{R}^n$.

(a) Show this product satisfies inner product laws 2,3 and 4 (page 267).

(b) Show that if A is a diagonal matrix with positive entries, then the product satisfies inner product law 1.

(c) Show that if A is a real symmetric positive definite matrix, then the product satisfies inner product law 1. Hint: Let P be an orthogonal matrix that diagonalizes A, write $\mathbf{x} = P\mathbf{y}$ and calculate $\langle \mathbf{x},\mathbf{x}\rangle$. Now use Exercise 15 and part (b).

12. Let $\mathbf{v}_1 = (1,0,0)$, $\mathbf{v}_2 = (-1,2,0)$, $\mathbf{v}_3 = (1,-2,3)$. Let $V = \mathbb{R}^3$ be an inner product space with inner product defined by the formula $\langle \mathbf{x},\mathbf{y}\rangle = \mathbf{x}^T A\mathbf{y}$, where

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

(a) Use Exercise 11 to show that the formula really does define an inner product.

(b) Verify that $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ form an orthogonal basis of V.

(c) Find the coordinates of $(1,2,-2)$ with respect to this basis by using the orthogonal coordinates theorem.

13. Let V be an inner product space. Use the definition of inner product to prove that for $\mathbf{u}, \mathbf{v} \in V$, and scalar a, the following are true.

(a) $\langle \mathbf{0}, \mathbf{v} \rangle = 0$

(b) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle \equiv \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

(c) $\langle a\mathbf{u}, \mathbf{v} \rangle = \overline{a} \langle \mathbf{u}, \mathbf{v} \rangle$

14. Prove the following generalization of the Pythagorean Theorem: If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are pairwise orthogonal vectors in the inner product space V and $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \ldots + \mathbf{v}_n$, then

$$||\mathbf{v}||^2 = ||\mathbf{v}_1||^2 + ||\mathbf{v}_2||^2 + \ldots + ||\mathbf{v}_n||^2$$

15. Show that $||\cdot||_1$ is not an induced norm on $\mathbb{R}^2$. *Hint:* See Example 6.2.15.

16. Let V be an inner product space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $||\cdot||$. Express $||\mathbf{u} + \mathbf{v}||^2$ and $||\mathbf{u} - \mathbf{v}||^2$ in terms of inner products and use this to prove the *polarization identity*

$$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{4} \left\{ ||\mathbf{u} + \mathbf{v}||^2 - ||\mathbf{u} + \mathbf{v}||^2 \right\}$$

(This identity shows that any inner product can be recovered from its induced norm.)

6.3. Gram-Schmidt Algorithm

We have seen that orthogonal bases have some very pleasant properties, such as the ease with which we can compute coordinates. Our goal in this section is very simple: given a subspace V of some inner product space and a basis $\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_n$ of V, to turn this basis into an orthogonal basis. This is exactly what the Gram-Schmidt algorithm is designed to do.

Description of the Algorithm

THEOREM 6.3.1. *Let $\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_n$ be a basis of the inner product space V. Define vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ recursively by the formula* Gram-Schmidt Algorithm

$$\mathbf{v}_k = \mathbf{w}_k - \frac{\langle \mathbf{v}_1, \mathbf{w}_k \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \frac{\langle \mathbf{v}_2, \mathbf{w}_k \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 - \ldots - \frac{\langle \mathbf{v}_{k-1}, \mathbf{w}_k \rangle}{\langle \mathbf{v}_{k-1}, \mathbf{v}_{k-1} \rangle}\mathbf{v}_{k-1}, \quad k = 1, \ldots, n.$$

Then we have

1. *The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_k$ form an orthogonal set.*
2. *For each index $k = 1, \ldots n$,*

$$\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_k\} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_k\}.$$

PROOF. In the case $k = 1$, we have that the single vector $\mathbf{v}_1 = \mathbf{w}_1$ is an orthogonal set and certainly $\operatorname{span}\{\mathbf{w}_1\} = \operatorname{span}\{\mathbf{v}_1\}$. Now suppose that for some index $k > 1$ we have shown that $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_{k-1}$ is an orthogonal set such that

$$\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_{k-1}\} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_{k-1}\}$$

Then it is true that $\langle \mathbf{v}_r, \mathbf{v}_s \rangle = 0$ for any indices r, s both less than k. Take the inner product of $\mathbf{v}_k$, as given by the formula above, with the vector $\mathbf{v}_j$, where $j < k$ and we obtain

$$\begin{aligned}
\langle \mathbf{v}_j, \mathbf{v}_k \rangle &= \left\langle \mathbf{v}_j, \mathbf{w}_k - \frac{\langle \mathbf{v}_1, \mathbf{w}_k \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \frac{\langle \mathbf{v}_2, \mathbf{w}_k \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 - \cdots - \frac{\langle \mathbf{v}_{k-1}, \mathbf{w}_k \rangle}{\langle \mathbf{v}_{k-1}, \mathbf{v}_{k-1} \rangle}\mathbf{v}_{k-1} \right\rangle \\
&= \langle \mathbf{v}_j, \mathbf{w}_k \rangle - \langle \mathbf{v}_1, \mathbf{w}_k \rangle \frac{\langle \mathbf{v}_j, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} - \cdots - \langle \mathbf{v}_{k-1}, \mathbf{w}_k \rangle \frac{\langle \mathbf{v}_j, \mathbf{v}_{k-1} \rangle}{\langle \mathbf{v}_{k-1}, \mathbf{v}_{k-1} \rangle} \\
&= \langle \mathbf{v}_j, \mathbf{w}_k \rangle - \langle \mathbf{v}_j, \mathbf{w}_k \rangle \frac{\langle \mathbf{v}_j, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle} \\
&= 0
\end{aligned}$$

It follows that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is an orthogonal set. The Gram-Schmidt formula show us that one of $\mathbf{v}_k$ or $\mathbf{w}_k$ can be expressed as a linear combination of the other and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}$. Therefore

$$\begin{aligned}
\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_{k-1}, \mathbf{w}_k\} &= \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_{k-1}, \mathbf{w}_k\} \\
&= \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_{k-1}, \mathbf{v}_k\}
\end{aligned}$$

which is the second part of the theorem. We can repeat this argument for each index $k = 2, \ldots, n$ to complete the proof of the theorem. $\square$

The Gram-Schmidt formula is easy to remember: One simply subtracts from the vector $\mathbf{w}_k$ all of the projections of $\mathbf{w}_k$ along the directions $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_{k-1}$ to obtain the vector $\mathbf{v}_k$. The Gram-Schmidt algorithm applies to any inner product space, not just the standard ones. Consider the following example.

EXAMPLE 6.3.2. Let $C[0, 1]$ be the space of continuous functions on the interval $[0, 1]$ with the usual function addition and scalar multiplication, and (standard) inner product given by

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx$$

as in Example 6.2.5. Let $V = \mathcal{P}_2 = \operatorname{span}\{1, x, x^2\}$ and apply the Gram-Schmidt algorithm to the basis $1, x, x^2$ to obtain an orthogonal basis for the space of quadratic polynomials.

SOLUTION. It helps to recall the calculus formula $\int_0^1 x^m x^n dx = 1/(m+n+1)$. Now set $\mathbf{w}_1 = 1, \mathbf{w}_2 = x, \mathbf{w}_3 = x^2$ and calculate the Gram-Schmidt formulas:

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{w}_1 = 1, \\
\mathbf{v}_2 &= \mathbf{w}_2 - \frac{\langle \mathbf{v}_1, \mathbf{w}_2 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 \\
&= x - \frac{1/2}{1} 1 = x - \frac{1}{2} \\
\mathbf{v}_3 &= \mathbf{w}_3 - \frac{\langle \mathbf{v}_1, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{v}_2, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 \\
&= x^2 - \frac{1/3}{1} 1 - \frac{1/12}{1/12} (x - \frac{1}{2}) \\
&= x^2 - x + \frac{1}{6}.
\end{aligned}$$

□

Had we used $C[-1, 1]$ and required that each polynomial have value 1 at $x = 1$, the same calculations would have given us the first three so-called *Legendre polynomials*. These polynomials are used extensively in approximation theory and applied mathematics.

If we prefer to have an orthonormal basis rather than an orthogonal basis, then, as a final step in the orthogonalizing process, simply replace each vector $\mathbf{v}_k$ by the normalized vector $\mathbf{u}_k = \mathbf{v}_k / ||\mathbf{v}_k||$. Here is an example to illustrate the whole scheme.

EXAMPLE 6.3.3. Let $V = \mathcal{C}(A)$, where

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

and V has the standard inner product. Find an orthonormal basis of V.

SOLUTION. We know that V is spanned by the four columns of A. However, the Gram-Schmidt algorithm requests a basis of V and we don't know that the columns are linearly independent. We leave it to the reader to check that the reduced row echelon form of A is the matrix

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

It follows from the column space algorithm that columns $1, 2$ and 4 of the matrix A yield a basis of V. So let $\mathbf{w}_1 = (1, 1, 1, -1), \mathbf{w}_2 = (2, -1, -1, 1), \mathbf{w}_3 = (-1, 2, 2, 1)$

and apply the Gram-Schmidt algorithm to obtain that

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{w}_1 = (1, 1, 1, -1), \\
\mathbf{v}_2 &= \mathbf{w}_2 - \frac{\langle \mathbf{v}_1, \mathbf{w}_2 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 \\
&= (2, -1, -1, 1) - \frac{-1}{4}(1, 1, 1, -1) \\
&= \frac{1}{4}(9, -3, -3, 3), \\
\mathbf{v}_3 &= \mathbf{w}_3 - \frac{\langle \mathbf{v}_1, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{v}_2, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 \\
&= (-1, 2, 2, 1) - \frac{2}{4}(1, 1, 1, -1) - \frac{-18}{108}(9, -3, -3, 3) \\
&= \frac{1}{4}(-4, 8, 8, 4) - \frac{1}{4}(2, 2, 2, -2) + \frac{1}{4}(6, -2, -2, 2) \\
&= (0, 1, 1, 2)\,.
\end{aligned}$$

Finally, to turn this set into an orthonormal basis, we normalize each vector to obtain the basis

$$\begin{aligned}
\mathbf{u}_1 &= \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{1}{2}(1, 1, 1, -1), \\
\mathbf{u}_2 &= \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{1}{\sqrt{108}}(9, -3, -3, 3) = \frac{1}{2\sqrt{3}}(3, -1, -1, 1), \\
\mathbf{u}_3 &= \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \frac{1}{\sqrt{6}}(0, 1, 1, 2).
\end{aligned}$$

□

There are several useful observations about the preceding example which are particularly helpful for hand calculations.

- If one encounters an inconvenient fraction, such as the $\frac{1}{4}$ in $\mathbf{v}_2$, one could replace the calculated $\mathbf{v}_2$ by $4\mathbf{v}_2$, thereby eliminating the fraction, and yet achieving the same results in subsequent calculations. The idea here is that for any nonzero scalar c
$$\frac{\langle \mathbf{v}_2, \mathbf{w} \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 = \frac{\langle c\mathbf{v}_2, \mathbf{w} \rangle}{\langle c\mathbf{v}_2, c\mathbf{v}_2 \rangle} c\mathbf{v}_2.$$
So we could have replaced $\frac{1}{4}(9, -3, -3, 3)$ by $(3, -1, -1, 1)$ and achieved the same results.
- The same remark applies to the normalizing process, since in general,
$$\frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{c\mathbf{v}_2}{\|c\mathbf{v}_2\|}$$

The Gram-Schmidt algorithm is robust enough to handle linearly dependent spanning sets gracefully. We illustrate this fact with the following example:

EXAMPLE 6.3.4. Suppose we had used all the columns of A in Example 6.3.3 instead of linearly independent ones, labelling them $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4$. How would the Gram-Schmidt calculation work out?

SOLUTION. Everything would have proceeded as above until we reached the calculation of $\mathbf{v}_3$, which would then yield

$$\begin{aligned}\mathbf{v}_3 &= \mathbf{w}_3 - \frac{\langle \mathbf{v}_1, \mathbf{w}_3\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle}\mathbf{v}_1 - \frac{\langle \mathbf{v}_2, \mathbf{w}_3\rangle}{\langle \mathbf{v}_2, \mathbf{v}_2\rangle}\mathbf{v}_2 \\ &= (0,3,3,-3) - \frac{9}{4}(1,1,1,-1) + \frac{1}{4}(9,-3,-3,3) \\ &= \frac{1}{4}(0,12,12,-12) + \frac{9}{4}(-1,-1,-1,1) - \frac{-27}{108}(9,-3,-3,3) \\ &= (0,0,0,0)\end{aligned}$$

This tells us that $\mathbf{v}_3$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$, which mirrors the fact that $\mathbf{w}_3$ is a linear combination of $\mathbf{w}_1$ and $\mathbf{w}_2$. Now discard $\mathbf{v}_3$ and continue the calculations to get that

$$\begin{aligned}\mathbf{v}_4 &= \mathbf{w}_4 - \frac{\langle \mathbf{v}_1, \mathbf{w}_4\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle}\mathbf{v}_1 - \frac{\langle \mathbf{v}_2, \mathbf{w}_4\rangle}{\langle \mathbf{v}_2, \mathbf{v}_2\rangle}\mathbf{v}_2 \\ &= (-1,2,2,1) - \frac{2}{4}(1,1,1,-1) - \frac{-18}{108}(9,-3,-3,3) \\ &= (0,1,1,2)\end{aligned}$$

Interestingly enough, this is the same third vector that we obtained in the example calculation. The upshot of this calculation is that the Gram-Schmidt algorithm can be applied to any spanning set, provided that one discards any zero vectors that result from the formula. The net result is still an orthogonal basis. $\square$

Application to Projections

We can use the machinery of orthogonal vectors to give a nice solution to a very practical and important question which can be phrased as follows (see Figure 6.3.1 for a graphical interpretation of it):

The Projection Problem: Given a finite dimensional subspace V of a real inner product space W, together with a vector $\mathbf{b} \in W$, to find the vector $\mathbf{v} \in V$ which is closest to $\mathbf{b}$ in the sense that $||\mathbf{b} - \mathbf{v}||^2$ is minimized.

Observe that the quantity $||\mathbf{b} - \mathbf{v}||^2$ will be minimized exactly when $||\mathbf{b} - \mathbf{v}||$ is minimized, since the latter is always nonnegative. The squared term has the virtue of avoiding square roots that computing $||\mathbf{b} - \mathbf{v}||$ requires.

The projection problem looks vaguely familiar. It reminds us of the least squares problem of Chapter 4, which was to minimize the quantity $||\mathbf{b} - A\mathbf{x}||^2$where A is an $m \times n$ real matrix and $\mathbf{b}, \mathbf{x}$ are standard vectors. Recall that $\mathbf{v} = A\mathbf{x}$ is a typical element in the column space of A. Therefore, the quantity to be minimized is

$$||\mathbf{b} - A\mathbf{x}||^2 = ||\mathbf{b} - \mathbf{v}||^2$$

where on the left hand side $\mathbf{x}$ runs over all standard n-vectors and on the right hand side $\mathbf{v}$ runs over all vectors in the space $V = \mathcal{C}(A)$. The difference between least squares and projection problem is this: in the least squares problem we want to know the vector $\mathbf{x}$ of coefficients of $\mathbf{v}$ as a linear combination of columns of A, whereas in the projection

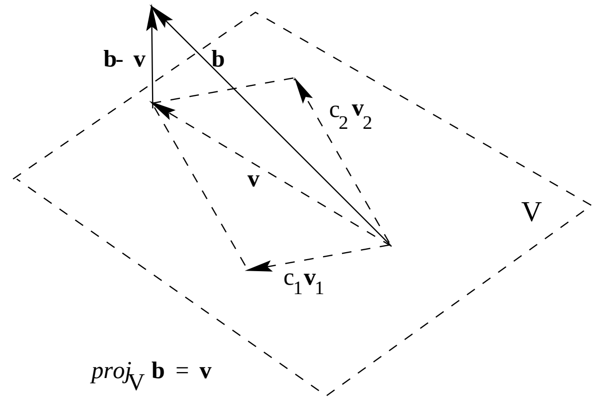


FIGURE 6.3.1. Projection $\mathbf{v}$ of $\mathbf{b}$ into subspace V spanned by orthogonal $\mathbf{v}_1, \mathbf{v}_2$.

problem we are only interested in $\mathbf{v}$. Knowing $\mathbf{v}$ doesn't tell us what $\mathbf{x}$ is, but knowing $\mathbf{x}$ easily gives $\mathbf{v}$ since $\mathbf{v} = A\mathbf{x}$.

To find a solution to the projection problem we need the following key concept.

Projection Formula for Subspaces

DEFINITION 6.3.5. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ be an orthogonal basis for the subspace V of the inner product space W. For any $\mathbf{b} \in \mathbf{W}$, the *projection of* $\mathbf{b}$ *into the subspace* V is the vector

$$\operatorname{proj}_V \mathbf{b} = \frac{\langle \mathbf{v}_1, \mathbf{b} \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{v}_2, \mathbf{b} \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{v}_n, \mathbf{b} \rangle}{\langle \mathbf{v}_n, \mathbf{v}_n \rangle}\mathbf{v}_n$$

Notice that in the case of $n = 1$ the definition amounts to a familiar friend, the projection of $\mathbf{b}$ along the vector $\mathbf{v}_1$. Now we call this the projection of $\mathbf{b}$ into the subspace V spanned by $\mathbf{v}_1$. This projection has the same nice property that we observed in the case of standard inner products, namely, $\mathbf{p} = \operatorname{proj}_V \mathbf{b}$ is a multiple of $\mathbf{v}_1$ which is orthogonal to $\mathbf{b} - \mathbf{p}$. Simply check that

$$\langle \mathbf{v}_1, \mathbf{b} - \mathbf{p} \rangle = \langle \mathbf{v}_1, \mathbf{b} \rangle - \left\langle \mathbf{v}_1, \frac{\langle \mathbf{v}_1, \mathbf{b} \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 \right\rangle = \langle \mathbf{v}_1, \mathbf{b} \rangle - \frac{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\langle \mathbf{v}_1, \mathbf{b} \rangle = 0.$$

It would appear that the definition depends on the basis vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$, but we see from the next theorem that this is not the case.

Projection Theorem

THEOREM 6.3.6. *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ be an orthogonal basis for the subspace V of the inner product space W. For any $\mathbf{b} \in \mathbf{W}$, the vector $\mathbf{v} = \operatorname{proj}_V \mathbf{b}$ is the unique vector in V that minimizes $||\mathbf{b} - \mathbf{v}||^2$.*

PROOF. Let $\mathbf{v}$ be a solution to the projection problem and $\mathbf{p}$ the projection of $\mathbf{b} - \mathbf{v}$ along any vector in V. Use the remark preceding this theorem with $\mathbf{b} - \mathbf{v}$ in place of $\mathbf{b}$ to write $\mathbf{b} - \mathbf{v}$ as the sum of orthogonal vectors $\mathbf{b} - \mathbf{v} - \mathbf{p}$ and $\mathbf{p}$. Now use the Pythagorean Theorem to see that

$$||\mathbf{b} - \mathbf{v}||^2 = ||\mathbf{b} - \mathbf{v} - \mathbf{p}||^2 + ||\mathbf{p}||^2$$

However, $\mathbf{v}+\mathbf{p}\in V$ so that $||\mathbf{b}-\mathbf{v}||$ cannot be the minimum distance $\mathbf{b}$ to a vector in V unless $||\mathbf{p}||=0$. It follows that $\mathbf{b}-\mathbf{v}$ is orthogonal to any vector in V. Now let $\mathbf{v}_1,\mathbf{v}_2,\ldots\mathbf{v}_n$ be an *orthogonal* basis of V and express the vector $\mathbf{v}$ in the form

$$\mathbf{v}=c_1\mathbf{v}_1+c_2\mathbf{v}_2+\cdots+c_n\mathbf{v}_n$$

Then for each $\mathbf{v}_k$ we must have

$$\begin{aligned}0&=\langle\mathbf{v}_k,\mathbf{b}-\mathbf{v}\rangle=\langle\mathbf{v}_k,\mathbf{b}-c_1\mathbf{v}_1-c_2\mathbf{v}_2-\cdots-c_n\mathbf{v}_n\rangle\\&=\langle\mathbf{v}_k,\mathbf{b}\rangle-c_1\langle\mathbf{v}_k,\mathbf{v}_1\rangle-c_2\langle\mathbf{v}_k,\mathbf{v}_2\rangle-\cdots c_n\langle\mathbf{v}_k,\mathbf{v}_n\rangle\\&=\langle\mathbf{v}_k,\mathbf{b}\rangle-c_k\langle\mathbf{v}_k,\mathbf{v}_k\rangle\end{aligned}$$

from which we deduce that $c_k=\langle\mathbf{v}_k,\mathbf{b}\rangle/\langle\mathbf{v}_k,\mathbf{v}_k\rangle$. It follows that

$$\mathbf{v}=\frac{\langle\mathbf{v}_1,\mathbf{b}\rangle}{\langle\mathbf{v}_1,\mathbf{v}_1\rangle}\mathbf{v}_1+\frac{\langle\mathbf{v}_2,\mathbf{b}\rangle}{\langle\mathbf{v}_2,\mathbf{v}_2\rangle}\mathbf{v}_2+\cdots+\frac{\langle\mathbf{v}_n,\mathbf{b}\rangle}{\langle\mathbf{v}_n,\mathbf{v}_n\rangle}\mathbf{v}_n=\operatorname*{proj}_V\mathbf{b}$$

This proves that there can be only one solution to the projection problem, namely the one given by the projection formula above.

There is one more point to be established, namely that $\operatorname{proj}_V\mathbf{b}$ actually solves the projection problem. This is left to the exercises. □

It is worth noting that in proving the preceding theorem, we showed that $\operatorname{proj}_V\mathbf{b}$ is orthogonal to every element of a basis of V and therefore to every element of V, since such elements are linear combinations of the basis elements.

Let us specialize to standard real vectors and inner products and take a closer look at the formula for the projection operator in the case that $\mathbf{v}_1,\mathbf{v}_2,\ldots\mathbf{v}_n$ is an orthonormal set. We then have $\langle\mathbf{v}_j,\mathbf{v}_j\rangle=1$, so

$$\begin{aligned}\operatorname{proj}_V\mathbf{b}&=\langle\mathbf{v}_1,\mathbf{b}\rangle\mathbf{v}_1+\langle\mathbf{v}_2,\mathbf{b}\rangle\mathbf{v}_2+\cdots+\langle\mathbf{v}_n,\mathbf{b}\rangle\mathbf{v}_n\\&=\left(\mathbf{v}_1^T\mathbf{b}\right)\mathbf{v}_1+\left(\mathbf{v}_2^T\mathbf{b}\right)\mathbf{v}_2+\cdots+\left(\mathbf{v}_n^T\mathbf{b}\right)\mathbf{v}_n\\&=\mathbf{v}_1\mathbf{v}_1^T\mathbf{b}+\mathbf{v}_2\mathbf{v}_2^T\mathbf{b}+\cdots+\mathbf{v}_n\mathbf{v}_n^T\mathbf{b}\\&=\left(\mathbf{v}_1\mathbf{v}_1^T+\mathbf{v}_2\mathbf{v}_2^T+\cdots+\mathbf{v}_n\mathbf{v}_n^T\right)\mathbf{b}\\&=P\mathbf{b}.\end{aligned}$$

Thus we have the following expression for the matrix P :

Orthogonal Projection Formula

$$P=\mathbf{v}_1\mathbf{v}_1^T+\mathbf{v}_2\mathbf{v}_2^T+\cdots+\mathbf{v}_n\mathbf{v}_n^T$$

The significance of this expression for projections in standard spaces over the reals with the standard inner product is as follows: computing the projection of a vector into a subspace amounts to no more than multiplying the vector by a matrix P which can be computed from V. Even in the case $n=1$ this fact gives us a new slant on projections:

$$\operatorname*{proj}_{\mathbf{v}}\mathbf{u}=(\mathbf{v}\mathbf{v}^T)\mathbf{u}$$

So here we have $P=\mathbf{v}\mathbf{v}^T$.

Projection Matrices

The general projection matrix P has some interesting properties. It is symmetric, i.e., $P^T=P$, and idempotent, i.e., $P^2=P$. Therefore, this notation is compatible with the definition of projection matrix introduced in earlier exercises (see Exercise 8). Symmetry follows from the fact that $\left(\mathbf{v}_k\mathbf{v}_k^T\right)^T = \mathbf{v}_k\mathbf{v}_k^T$. For idempotence, notice that

$$(\mathbf{v}_j\mathbf{v}_j^T)(\mathbf{v}_k\mathbf{v}_k^T) = (\mathbf{v}_j^T\mathbf{v}_k)(\mathbf{v}_k\mathbf{v}_j^T) = \delta_{i,j}.$$

It follows that $P^2 = P$. In general, symmetric idempotent matrices are called *projection* matrices. The name is justified because multiplication by them projects vectors into the column space of P.

EXAMPLE 6.3.7. Find the projection matrix for the subspace of $\mathbb{R}^3$ spanned by the orthonormal vectors $\mathbf{v}_1 = (1/\sqrt{2})[1,-1,0]^T$ and $\mathbf{v}_2 = (1/\sqrt{3})[1,1,1]^T$ and use it to solve the projection problem with $V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and $\mathbf{b} = [2,1,-3]^T$.

SOLUTION. Use the formula developed above for the projection matrix

$$\begin{aligned} P &= \mathbf{v}_1\mathbf{v}_1^T + \mathbf{v}_2\mathbf{v}_2^T \\ &= \frac{1}{2}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}\begin{bmatrix} 1 & -1 & 0 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \\ &\begin{bmatrix} \frac{5}{6} & -\frac{1}{6} & \frac{1}{3} \\ -\frac{1}{6} & \frac{5}{6} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{aligned}$$

The solution to the projection problem is now given by

$$\mathbf{v} = P\mathbf{b} = \begin{bmatrix} \frac{5}{6} & -\frac{1}{6} & \frac{1}{3} \\ -\frac{1}{6} & \frac{5}{6} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 0 \end{bmatrix}$$

□

The projection problem is closely related to another problem that we have seen before, namely the least squares problem of Section 4.2 in Chapter 4. Recall that the least squares problem amounted to minimizing the function $f(x) = ||\mathbf{b} - A\mathbf{x}||^2$, which in turn led to the normal equations. Here A is an $m \times n$ real matrix. Now consider the projection problem for the subspace $V = \mathcal{C}(A)$ of $\mathbb{R}^m$, where $\mathbf{b} \in \mathbb{R}^m$. We know that elements of $\mathcal{C}(A)$ can be written in the form $\mathbf{v} = A\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^n$. Therefore, $||\mathbf{b} - A\mathbf{x}||^2 = ||\mathbf{b} - \mathbf{v}||^2$, where $\mathbf{v}$ ranges over elements of V. It follows that when we solve a least squares problem, we are solving a projection problem as well in the sense that the vector $A\mathbf{x}$ is the element of $\mathcal{C}(A)$ closest to the right hand side vector $\mathbf{b}$. One could also develop normal equations for general spanning sets of V. An example of this is given in the exercises.

The normal equations also give us another way to generate projection matrices in the case of standard vectors and inner products. As above, let the subspace $V = \mathcal{C}(A)$ of $\mathbb{R}^m$, and $\mathbf{b} \in \mathbb{R}^m$. Assume that $V = \mathcal{C}(A)$ and that the columns of A are linearly independent, i.e., that A has full column rank. Then, as we have seen in Theorem 4.2.11 of Chapter 3, the matrix A^TA is invertible and the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$ have the unique solution

$$\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b}.$$

Consequently, the solution to the projection problem is

$$\mathbf{v} = A\mathbf{x} = A(A^T A)^{-1} A^T \mathbf{b}.$$

It is also true that $\mathbf{v} = P\mathbf{b}$; since this holds for all $\mathbf{b}$, it follows that the projection matrix for this subspace is given by the formula

Column Space Projection Formula

$$P = A(A^T A)^{-1} A^T$$

EXAMPLE 6.3.8. Find the projection matrix for the subspace $V = \operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2\}$ of $\mathbb{R}^3$ with $\mathbf{w}_1 = (1, -1, 0)$ and $\mathbf{w}_2 = (2, 0, 1)$.

SOLUTION. Let $A = [\mathbf{w}_1, \mathbf{w}_2]$ so that

$$A^T A = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}$$

Thus

$$\begin{aligned} P &= A(A^T A)^{-1} A^T \\ &= \begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{6} \begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{5}{6} & -\frac{1}{6} & \frac{1}{3} \\ -\frac{1}{6} & \frac{5}{6} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{aligned}$$

$\square$

Curiously, this is exactly the same matrix as the projection matrix found in the preceding example. What is the explanation? Notice that $\mathbf{w}_1 = \sqrt{6}\mathbf{v}_1$ and $\mathbf{w}_2 = \sqrt{6}\mathbf{v}_1 + \sqrt{3}\mathbf{v}_2$, so that $V = \operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2\} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$. So the subspace of both examples is the same, but specified by different bases. Therefore we should expect the projection operator to be the same.

6.3 Exercises

1. Find the projection of the vector $\mathbf{w} = (2, 1, 2)$ into the subspace

$$V = \operatorname{span}\{(1, -1, 1), (1, 1, 0)\}$$

where the inner products used are the standard inner product on $\mathbb{R}^3$ and the weighted inner product

$$< (x, y, z), (u, v, w) >= 2xu + 3yv + zw$$

2. Let $[\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3] = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -1 \\ 0 & -1 & 1 \end{bmatrix}$ and $\mathbf{w} = (2, 1, 4)$.

(a) Use the Gram-Schmidt algorithm on $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ to obtain an orthonormal set $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$.

(b) Find the projection of $\mathbf{w}$ into the subspace $V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$.

(c) Use (b) to express $\mathbf{w}$ as a sum of orthogonal vectors, one of which is in V.

(d) Find the projection matrix P for the subspace V.

(e) Verify that multiplication of $\mathbf{w}$ by P gives the same result as in (b).

3. Let $\mathbf{w}_1 = (-1,-1,1,1), \quad \mathbf{w}_2 = (1,1,1,1), \quad \mathbf{w}_3 = (0,0,0,1), \quad \mathbf{w} = (1,0,0,0)$ and $V = \operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$. Repeat parts (a)-(e) of Exercise 2 for these vectors, except in part (b) use $V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

4. Find an orthonormal basis of $\mathcal{C}(A)$, where A is one of the following

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

5. Find the projection matrix for the column space of each of the following matrices by using the orthogonal projection formula.

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

6. Redo Exercise 5 by using the column space projection formula (remember to use a matrix of full column rank for this formula).

7. Show that the matrices $A = \begin{bmatrix} 1 & 3 & 4 \\ 1 & 4 & 2 \\ 1 & 1 & 8 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 2 \\ -2 & -3 & -2 \\ 7 & 12 & 10 \end{bmatrix}$ have the same column space by computing the projection matrices into these column spaces.

8. Let $W = C[-1,1]$ with the standard inner product as in Example 6.3.2. Suppose V is the subspace of linear polynomials and $\mathbf{b} = e^x$.

(a) Find an orthogonal basis for V.

(b) Find the projection $\mathbf{p}$ of $\mathbf{b}$ into V.

(c) Compute the "mean error of approximation" $||\mathbf{b} - \mathbf{p}||$. How does it compare to the mean error of approximation when one approximates $\mathbf{b}$ by $\mathbf{q}$, its Taylor series centered at 0.

(d) Use a CAS to plot $\mathbf{b} - \mathbf{p}$ and $\mathbf{b} - \mathbf{q}$. Find the points at which this error is largest and compare the two.

9. Write out a proof of the Gram-Schmidt algorithm (Theorem 6.3.1) in the case that $n = 3$.

10. Complete the proof of the Projection Theorem (Theorem 6.3.6) by showing that $\operatorname{proj}_V \mathbf{b}$ solves the projection problem.

11. Verify directly that if $P = A(A^TA)^{-1}A^T$, (assume A has full column rank) then P is symmetric and idempotent.

12. How does the orthogonal projection formula on page 282 have to be changed if the vectors in question are complex? Illustrate your answer with the orthonormal vectors $\mathbf{v}_1 = ((1+i)/2, 0, (1+i)/2), \mathbf{v}_2 = (0,1,0)$ in $\mathbb{C}^2$.

13. Show that if P is a square $n \times n$ real matrix such that $P^T = P$ and $P^2 = P$, that is, A is a projection matrix, then for every $\mathbf{v} \in \mathbb{R}^n$, $P\mathbf{v} \in \mathcal{C}(A)$ and $\mathbf{v} - P\mathbf{v}$ is orthogonal to every element of $\mathcal{C}(A)$.

6.4. Linear Systems Revisited

Once again, we revisit our old friend, $A\mathbf{x} = \mathbf{b}$, where A is an $m \times n$ matrix. The notions of orthogonality can be made to shed still more light on the nature of this system of equations, especially in the case of a homogeneous system $A\mathbf{x} = 0$. The kth entry of the column vector $A\mathbf{x}$ is simply the kth row of A multiplied by the column vector $\mathbf{x}$. Designate this row by r_k and we see that

$$\mathbf{r}_k \cdot \mathbf{x} = 0, \qquad k = 1, \ldots, n.$$

In other words, $A\mathbf{x} = 0$, that is, $\mathbf{x} \in \mathcal{N}(A)$, precisely when $\mathbf{x}$ is orthogonal (with the standard inner product) to every row of A. We will see in Theorem 6.4.4 below that this means that $\mathbf{x}$ will be orthogonal to any linear combination of the rows of A. Thus, we could say

$$\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{r} \cdot \mathbf{x} = 0 \text{ for every } \mathbf{r} \in \mathcal{R}(A)\} \tag{6.4.1}$$

This is an instance of a very important idea. We are going to digress to put this idea in a more general context, after which we will return to linear systems with a new perspective on their meaning.

Orthogonal Complements and Homogeneous Systems

DEFINITION 6.4.1. Let V be a subspace of an inner product space W. Then the *orthogonal complement* of V in W is the set

$$V^\perp = \{\mathbf{w} \in W \mid \langle \mathbf{v}, \mathbf{w} \rangle = 0 \quad \text{for all } \mathbf{v} \in V\}$$

We can see from the subspace test that $V^\perp$ is a subspace of W. Before stating the basic facts, we mention that if U and V are two subspaces of the vector space W, then two other subspaces that we can construct are the *intersection* and *sum* of these subspaces. The former is just the set intersection of the two subspaces and the latter is the set of elements of the form $\mathbf{u} + \mathbf{v}$, where $\mathbf{u} \in U$, and $\mathbf{v} \in V$. One can use the subspace test to verify that these are indeed subspaces of W (see Exercise 15 of Section 2, Chapter 3. In fact, it isn't too hard to see that $U + V$ is the smallest space containing all elements of both U and V. We can summarize the basic facts about the orthogonal complement of V as follows.

THEOREM 6.4.2. *Let V be a subspace of the finite dimensional inner product space W. Then the following are true:*

1. $V^{\perp}$ *is a subspace of* W.
2. $V \cap V^{\perp} = \{0\}$
3. $V + V^{\perp} = W$
4. $\dim V + \dim V^{\perp} = \dim W$
5. $\left(V^{\perp}\right)^{\perp} = V$

PROOF. We leave 1 and 2 as exercises. To prove 3, we notice that $V + V^{\perp} \subseteq W$ since W is closed under sums. Now suppose that $\mathbf{w} \in W$. Let $\mathbf{v} = \operatorname{proj}_V \mathbf{w}$. We know that $\mathbf{v} \in V$ and $\mathbf{w} - \mathbf{v}$ is orthogonal to every element of V. It follows that $\mathbf{w} - \mathbf{v} \in V^{\perp}$. Therefore every element of W can be expressed as a sum of an element in V and an element in $V^{\perp}$. This shows that $W \subseteq V + V^{\perp}$, from which it follows that $V + V^{\perp} = W$.

To prove 4, let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$ be a basis of V and $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_s$ be a basis of $V^{\perp}$. Certainly the union of the two sets spans V because of 3. Now if there were an equation of linear dependence, we could gather all terms involving $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$ on one side of the equation, those involving $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_s$ on the other side and deduce that each is equal to zero separately, in view of 2. It follows that the union of these two bases must be an independent set. Therefore it forms a basis of W. It follows that $\dim W = r + s = \dim V + \dim V^{\perp}$.

Finally, apply 4 to $V^{\perp}$ in place of V and obtain that $\dim \left(V^{\perp}\right)^{\perp} = \dim W - \dim V^{\perp}$. But 4 implies directly that $\dim V = \dim W - \dim V^{\perp}$, so that $\dim \left(V^{\perp}\right)^{\perp} = \dim V$. Now if $\mathbf{v} \in V$, then certainly $\langle \mathbf{w}, \mathbf{v} \rangle = 0$ for all $\mathbf{w} \in V^{\perp}$. Hence $V \subseteq \left(V^{\perp}\right)^{\perp}$. Since these two spaces have the same dimension, they must be equal, which proves 5. $\square$

Orthogonal complements of the sum and intersections of two different subspaces have an interesting relationship to each other. We will leave the proofs of these facts as exercises.

THEOREM 6.4.3. *Let U and V be subspaces of the inner product space W. Then the following are true:*

1. $(U \cap V)^{\perp} = U^{\perp} + V^{\perp}$
2. $(U + V)^{\perp} = U^{\perp} \cap V^{\perp}$

There is a very useful fact about the orthogonal complement of a finite dimensional space that greatly simplifies the calculation of an orthogonal complement. What it says in words is that a vector is orthogonal to every element of a vector space if and only if it is orthogonal to every element of a spanning set of the space.

THEOREM 6.4.4. *Let $V = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be a subspace of the inner product space W. Then*

$$V^{\perp} = \{\mathbf{w} \in W \mid \langle \mathbf{w}, \mathbf{v}_j \rangle = 0,\ j = 1, 2, \dots, n\}$$

PROOF. Let $\mathbf{v} \in V$, so that for some scalars $c_1, c_2, \dots, c_n$

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

Take the inner product of both sides with a vector $\mathbf{w}$. We see by the linearity of inner products that

$$\langle \mathbf{w}, \mathbf{v} \rangle = c_1\langle \mathbf{w}, \mathbf{v}_1 \rangle + c_2\langle \mathbf{w}, \mathbf{v}_2 \rangle + \cdots + c_n\langle \mathbf{w}, \mathbf{v}_n \rangle$$

so that if $\langle \mathbf{w}, \mathbf{v}_j \rangle = 0$ for each j then certainly $\langle \mathbf{w}, \mathbf{v} \rangle = 0$. Conversely, if $\langle \mathbf{w}, \mathbf{v}_j \rangle = 0$, $j = 1, 2, \ldots, n$, clearly $\langle \mathbf{w}, \mathbf{v}_j \rangle = 0$. This proves the theorem. □

EXAMPLE 6.4.5. Compute $V^{\perp}$, where

$$V = \operatorname{span}\{(1,1,1,1),(1,2,1,0)\} \subseteq \mathbb{R}^4$$

with the standard inner product on $\mathbb{R}^4$.

SOLUTION. Form the matrix A with the two spanning vectors of V as rows. According to Theorem 6.4.4, $V^{\perp}$ is simply the null space of this matrix. We have

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \end{bmatrix} \xrightarrow[E_{21}(-1)]{} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \xrightarrow[E_{12}(-1)]{} \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

from which it follows that the null space of A consists of vectors of the form

$$\begin{bmatrix} -x_3 - 2x_4 \\ x_4 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Therefore the null space is spanned by $(-1, 0, 1, 0)$ and $(-2, 1, 0, 1)$.

Nothing prevents us from considering more exotic inner products as well. The arithmetic may be a bit more complicated, but the underlying principles are the same. Here is such an example.

EXAMPLE 6.4.6. Let $V = \operatorname{span}\{1, x\} \subset W = \mathcal{P}_2$, where the space $\mathcal{P}_2$ of polynomials of degree at most 2 has the same standard inner product as $C[0,1]$. Compute $V^{\perp}$ and use this to verify that $\dim V + \dim V^{\perp} = \dim W$.

SOLUTION. According to Theorem 6.4.4, $V^{\perp}$ consists of those polynomials $p(x) = c_0 + c_1 x + c_2 x^2$ for which

$$0 = \langle p, 1 \rangle = \int_0^1 \left(c_0 + c_1 x + c_2 x^2\right) 1\, dx = c_0 \int_0^1 1\, dx + c_1 \int_0^1 x\ dx + c_2 \int_0^1 x^2\, dx$$

and

$$0 = \langle p, x \rangle = \int_0^1 \left(c_0 + c_1 x + c_2 x^2\right) x\, dx = c_0 \int_0^1 x\, dx + c_1 \int_0^1 x^2\ dx + c_2 \int_0^1 x^3\, dx$$

Use the fact that $\int_0^1 x^m\, dx = \frac{1}{m+1}$ for nonnegative m and we obtain the system of equations

$$\begin{aligned} c_0 + \frac{1}{2} c_1 + \frac{1}{3} c_2 &= 0 \\ \frac{1}{2} c_0 + \frac{1}{3} c_1 + \frac{1}{4} c_2 &= 0. \end{aligned}$$

Solve this system to obtain $c_0 = \frac{1}{6} c_2$, $c_1 = -c_2$ and c_2 is free. Therefore, $V^{\perp}$ consists of polynomials of the form

$$p(x) = \frac{1}{6} c_2 - c_2 x + c_2 x^2 = c_2 \left(\frac{1}{6} - x + x^2\right)$$

It follows that $V^\perp = \operatorname{span}\{\frac{1}{6} - x + x^2\}$. In particular, $\dim V^\perp = 1$, and since $\{1, x\}$ is a linearly independent set, $\dim V = 2$. Therefore, $\dim V + \dim V^\perp = \dim \mathcal{P}_2 = \dim W$. □

Finally, we return to the subject of solutions to the homogeneous system $A\mathbf{x} = \mathbf{b}$. We saw at the beginning of this section that the null space of A consisted of elements that are orthogonal to the rows of A. One could turn things around and ask what we can say about a vector that is orthogonal to every element of the null space of A. How does it relate to the rows of A? This question has a surprisingly simple answer. In fact, there is a fascinating interplay between row spaces, column spaces and null spaces which can be summarized in the following theorem:

Orthogonal Complements Theorem

THEOREM 6.4.7. *Let A be a matrix. Then*

1. $\mathcal{R}(A)^\perp = \mathcal{N}(A)$
2. $\mathcal{N}(A)^\perp = \mathcal{R}(A)$
3. $\mathcal{N}(A^T)^\perp = \mathcal{C}(A)$

PROOF. We have already seen item 1 in the discussion at the beginning of this section, where it was stated in Equation 6.4.1. For item 2 we take orthogonal complements of both sides of 1 and use part 5 of Theorem 6.4.2 to obtain that

$$\mathcal{N}(A)^\perp = \left(\mathcal{R}(A)^\perp\right)^\perp = \mathcal{R}(A)$$

which proves 2. Finally, for 3 we observe that $\mathcal{R}(A^T) = \mathcal{C}(A)$. Apply 2 with A^T in place of A and the result follows. □

The connections spelled out by this theorem are powerful ideas. Here is one example of how they can be used. Consider the following problem: suppose we are given subspaces U and V of the standard space $\mathbb{R}^n$ with the standard inner product (the dot product) in some concrete form, and we want to compute a basis for the subspace $U \cap V$. How do we proceed? One answer is to use part 1 of Theorem 6.4.3 to see that $(U \cap V)^\perp = U^\perp + V^\perp$. Now use part 5 of Theorem 6.4.2 to obtain that

$$U \cap V = (U \cap V)^{\perp\perp} = (U^\perp + V^\perp)^\perp$$

The strategy that this equation suggests is as follows: express U and V as row spaces of matrices and compute bases for the null spaces of each. Put these bases together to obtain a spanning set for $U^\perp + V^\perp$. Use this spanning set as the rows of a matrix B. Then the complement of this space is, on the one hand, $U \cap V$, but by the first part of the orthogonal complements theorem, it is also $\mathcal{N}(B)$. Therefore $U \cap V = \mathcal{N}(B)$, so all we have to do is calculate a basis for $\mathcal{N}(B)$, which we know how to do.

EXAMPLE 6.4.8. Find a basis for $U \cap V$, where these subspaces of $\mathbb{R}^4$ are given as follows:

$$\begin{aligned} U &= \operatorname{span}\{(1,2,1,2),(0,1,0,1)\} \\ V &= \operatorname{span}\{(1,1,1,1),(1,2,1,0)\}. \end{aligned}$$

SOLUTION. We have already determined in Exercise 6.4.5 that the null space of V has a basis $(-1,0,1,0)$ and $(-2,1,0,1)$. Similarly, form the matrix A with the two spanning vectors of U as rows. By Theorem 6.4.4, $V^\perp = \mathcal{N}(A)$. We have

$$A = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & 1 & 0 & 1 \end{bmatrix} \xrightarrow[E_{21}(-2)]{} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

from which it follows that the null space of A consists of vectors of the form

$$\begin{bmatrix} -x_3 \\ -x_4 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

Therefore the null space has basis $(-1,0,1,0)$ and $(0,1,0,1)$. One of the vectors in this set is repeated in the basis of V so we only to need list it once. Form the matrix B whose rows are $(-1,0,1,0)$, $(-2,1,0,1)$ and $(0,1,0,1)$, and calculate the reduced row echelon form of B :

$$B = \begin{bmatrix} -1 & 0 & 1 & 0 \\ -2 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \xrightarrow[\substack{E_{12}(-2) \\ E_1(-1)}]{} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow[E_{23}(-1)]{} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 2 & 0 \end{bmatrix} \xrightarrow[\substack{E_3(1/2) \\ E_{32}(2) \\ E_{31}(1)}]{} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

It follows that the null space of B consists of vectors of the form

$$\begin{bmatrix} 0 \\ -x_4 \\ 0 \\ x_4 \end{bmatrix} = x_4 \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}.$$

Therefore, $U \cap V$ is a one-dimensional space spanned by the vector $(0,-1,0,1)$. □

Our last application of the orthogonal complements theorem is another Fredholm alternative theorem (compare this to Corollary 2.5.10 of Chapter 2).

COROLLARY 6.4.9. *Given a square real linear system $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} \neq 0$, either the system is consistent or there is a solution $\mathbf{y}$ to the homogeneous system $A^T\mathbf{y} = 0$ such that $\mathbf{y}^T\mathbf{b} \neq 0$.* Fredholm Alternative

PROOF. Let $V = \mathcal{C}(A)$. By part 3 of Theorem 6.4.2, $\mathbb{R}^n = V + V^\perp$, where $\mathbb{R}^n$ has the standard inner product. From part 3 of the orthogonal complements theorem, $\mathcal{C}(A) = \mathcal{N}(A^T)^\perp$. Take complements again and use part 5 of Theorem 6.4.2 to get that $V^\perp = \mathcal{N}(A^T)$. Now the system either has a solution or not. If the system has no solution, then by Theorem 3.6.1 of Chapter 3, $\mathbf{b}$ does not belong to $V = \mathcal{C}(A)$. Since $\mathbf{b} \notin V$, we can write $\mathbf{b} = \mathbf{v} + \mathbf{y}$, where $\mathbf{y} \neq 0$, $\mathbf{y} \in V^\perp$ and $\mathbf{v} \in V$. It follows that

$$\langle \mathbf{y}, \mathbf{b} \rangle = \mathbf{y} \cdot \mathbf{b} = \mathbf{y} \cdot (\mathbf{v} + \mathbf{y}) = 0 + \mathbf{y} \cdot \mathbf{y} \neq 0$$

On the other hand, if the system has a solution $\mathbf{x}$, then for any vector $\mathbf{y} \in \mathcal{N}(A)$ we have $\mathbf{y}^T A\mathbf{x} = \mathbf{y}^T\mathbf{b}$. It follows that if $\mathbf{y}^T A = 0$, then $\mathbf{y}^T\mathbf{b} = 0$. This completes the proof. □

The QR Factorization

We are going to use orthogonality ideas to develop one more way of solving the linear system $A\mathbf{x} = \mathbf{b}$, where the $m \times n$ real matrix A is full column rank. In fact, if the system is inconsistent, then this method will find the unique least squares solution to the system. Here is the basic idea: express the matrix A in the form $A = QR$, where the columns of the $m \times n$ matrix Q are orthonormal vectors and the $n \times n$ matrix R is upper triangular with nonzero diagonal entries. Such a factorization of A is called a ***QR factorization*** of A. It follows that the product $Q^T Q = I_n$. Now multiply both sides of the linear system on the left by Q^T to obtain that

$$Q^T A\mathbf{x} = Q^T QR\mathbf{x} = IR\mathbf{x} = Q^T b$$

The net result is a simple square system with a triangular matrix which we can solve by backsolving. That is, we use the last equation to solve for x_n, then the next to the last to solve for x_{n-1}, and so forth. This is the backsolving phase of Gaussian elimination as we first learned it in Chapter 1, before we were introduced to Gauss-Jordan elimination.

One has to wonder why we have any interest in such a factorization, since we already have Gauss-Jordan elimination for system solving. Furthermore, it can be shown that finding a QR factorization is harder by a factor of about 2, that is, requires about twice as many floating point operations to accomplish. So why bother? There are many answers. For one, it can be shown that using the QR factorization has an advantage of higher accuracy than Gauss-Jordan elimination in certain situations. For another, QR factorization gives us another method for solving least squares problems. We'll see an example of this method at the end of this section.

Where can we find such a factorization? As a matter of fact, we already have the necessary tools, compliments of the Gram-Schmidt algorithm. To explain matters, let's suppose that we have a matrix $A = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3]$ with linearly independent columns. Application of the Gram-Schmidt algorithm leads to orthogonal vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ by the following formulas:

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{w}_1 \\
\mathbf{v}_2 &= \mathbf{w}_2 - \frac{\langle \mathbf{v}_1, \mathbf{w}_2 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{w}_3 - \frac{\langle \mathbf{v}_1, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - - \frac{\langle \mathbf{v}_2, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2
\end{aligned}$$

Next, solve for $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ in the above equations to obtain

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v}_1 \\
\mathbf{w}_2 &= \frac{\langle \mathbf{v}_1, \mathbf{w}_2 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \mathbf{v}_2 \\
\mathbf{w}_3 &= \frac{\langle \mathbf{v}_1, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{v}_2, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 + \mathbf{v}_3
\end{aligned}$$

In matrix form, these equations become

$$A = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3] = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] \begin{bmatrix} 1 & \frac{\langle \mathbf{v}_1, \mathbf{w}_2 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} & \frac{\langle \mathbf{v}_1, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \\ 0 & 1 & \frac{\langle \mathbf{v}_2, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \\ 0 & 0 & 1 \end{bmatrix}$$

Now normalize the $\mathbf{v}_j$'s by setting $\mathbf{q}_j = \mathbf{v}_j / \|\mathbf{v}_j\|$ and observe that

$$\begin{aligned} A &= [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3] \begin{bmatrix} \|\mathbf{v}_1\| & 0 & 0 \\ 0 & \|\mathbf{v}_2\| & 0 \\ 0 & 0 & \|\mathbf{v}_3\| \end{bmatrix} \begin{bmatrix} 1 & \frac{\langle \mathbf{v}_1, \mathbf{w}_2 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} & \frac{\langle \mathbf{v}_1, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \\ 0 & 1 & \frac{\langle \mathbf{v}_2, \mathbf{w}_3 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \\ 0 & 0 & 1 \end{bmatrix} \\ &= [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3] \begin{bmatrix} \|\mathbf{v}_1\| & \frac{\langle \mathbf{v}_1, \mathbf{w}_2 \rangle}{\|\mathbf{v}_1\|} & \frac{\langle \mathbf{v}_1, \mathbf{w}_3 \rangle}{\|\mathbf{v}_1\|} \\ 0 & \|\mathbf{v}_2\| & \frac{\langle \mathbf{v}_2, \mathbf{w}_3 \rangle}{\|\mathbf{v}_2\|} \\ 0 & 0 & \|\mathbf{v}_3\| \end{bmatrix} \end{aligned}$$

This gives our QR factorization, which can be alternately written as

$$A = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3] = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3] \begin{bmatrix} \|\mathbf{v}_1\| & \langle \mathbf{q}_1, \mathbf{w}_2 \rangle & \langle \mathbf{q}_1, \mathbf{w}_3 \rangle \\ 0 & \|\mathbf{v}_2\| & \langle \mathbf{q}_2, \mathbf{w}_3 \rangle \\ 0 & 0 & \|\mathbf{v}_3\| \end{bmatrix} = QR$$

In general, the columns of A are linearly independent exactly when A is full column rank. It is easy to see that the argument we have given extends to any such matrix, so we have the following theorem.

THEOREM 6.4.10. *If A is an $m \times n$ matrix full column rank matrix, then $A = QR$, where the columns of the $m \times n$ matrix Q are orthonormal vectors and the $n \times n$ matrix R is upper triangular with nonzero diagonal entries.* QR Factorization

EXAMPLE 6.4.11. Let the full column rank matrix A be given as

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

Find a QR factorization of A and use this to find the least squares solution to the problem $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = (1, 1, 1, 1)$. What is the norm of the residual $\mathbf{r} = \mathbf{b} - A\mathbf{x}$ in this problem?

SOLUTION. Notice that the columns of A are just the vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ of Example 6.3.3. Furthermore, the vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ calculated in that example are just the $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ that we require. Thus we have from those calculations that

$$\begin{aligned} \|\mathbf{v}_1\| &= \|(1, 1, 1, -1)\| = 2 \\ \|\mathbf{v}_2\| &= \left\| \frac{1}{4}(9, -3, -3, 3) \right\| = \frac{3}{2}\sqrt{3} \\ \|\mathbf{v}_3\| &= \|(0, 1, 1, 2)\| = \sqrt{6} \\ \mathbf{q}_1 &= \frac{1}{2}(1, 1, 1, -1) \\ \mathbf{q}_2 &= \frac{1}{2\sqrt{3}}(3, -1, -1, 1) \\ \mathbf{q}_3 &= \frac{1}{\sqrt{6}}(0, 1, 1, 2) \end{aligned}$$

Now we calculate

$$\begin{aligned}
\langle \mathbf{q}_1, \mathbf{w}_2\rangle &= \frac{1}{2}(1,1,1,-1)\cdot(2,-1,-1,1) = -\frac{1}{2}\\
\langle \mathbf{q}_1, \mathbf{w}_3\rangle &= \frac{1}{2}(1,1,1,-1)\cdot(-1,2,2,1) = 1\\
\langle \mathbf{q}_2, \mathbf{w}_3\rangle &= \frac{1}{2\sqrt{3}}(3,-1,-1,1)\cdot(-1,2,2,1) = -\sqrt{3}
\end{aligned}$$

It follows that

$$A = \begin{bmatrix} 1/2 & 3/(2\sqrt{3}) & 0\\ 1/2 & -1/(2\sqrt{3}) & 1/\sqrt{6}\\ 1/2 & -1/(2\sqrt{3}) & 1/\sqrt{6}\\ -1/2 & 1/(2\sqrt{3}) & 2/\sqrt{6}\end{bmatrix}\begin{bmatrix} 2 & -1/2 & 1\\ 0 & \frac{3}{2}\sqrt{3} & -\sqrt{3}\\ 0 & 0 & \sqrt{6}\end{bmatrix} = QR$$

Solving the system $R\mathbf{x} = Q^T\mathbf{b}$, where $\mathbf{b} = (1,1,1,1)$, by hand is rather tedious even though the system is a simple triangular one. We leave the detailed calculations to the reader. Better yet, use a CAS or MAS to obtain the solution $\mathbf{x} = \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)$. Another calculation shows that

$$\mathbf{r} = \mathbf{b} - A\mathbf{x} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} - \begin{bmatrix} 1 & 2 & -1\\ 1 & -1 & 2\\ 1 & -1 & 2\\ -1 & 1 & 1\end{bmatrix}\begin{bmatrix}1/3\\2/3\\2/3\end{bmatrix} = \begin{bmatrix}0\\0\\0\\0\end{bmatrix}$$

It follows that the system $A\mathbf{x} = \mathbf{b}$ is actually consistent, since the least squares solution turns out to be a genuine solution to the problem. □

QR Least Square Solver

There remains the question of why we really solve the least squares problem by this method. To see why this is so, notice that with the above notation we have $A^T = (QR)^T = R^TQ^T$, so that the normal equations for the system $A\mathbf{x} = \mathbf{b}$ (which are given by $A^TA\mathbf{x} = A^T\mathbf{b}$) become

$$A^TA\mathbf{x} = R^TQ^TQR\mathbf{x} = R^TIR\mathbf{x} = R^TR\mathbf{x} = A^T\mathbf{b} = R^TQ^T\mathbf{b}.$$

But the triangular matrix R is invertible because its diagonal entries are nonzero; cancel it and obtain that the normal equations are equivalent to $R\mathbf{x} = Q^T\mathbf{b}$, which is exactly what the method we have described solves.

6.4 Exercises

1. Let $V = \operatorname{span}\{(1,-1,2)\} \subset \mathbb{R}^3 = W$ with the standard inner product.

(a) Compute $V^\perp$.

(b) Verify that $V + V^\perp = \mathbb{R}^3$ and $V \cap V^\perp = \{0\}$.

2. Let $V = \operatorname{span}\{1+x, x^2\} \subset W = \mathcal{P}_2$, where the space $\mathcal{P}_2$ of polynomials of degree at most 2 has the same standard inner product as $C[0,1]$. Compute $V^\perp$.

3. Let V be as in Exercise 1 but endow W with the weighted inner product $<(x,y,z),(u,v,w)>= 2xu+3yv+zw$.

(a) Compute $V^{\perp}$.

(b) Verify that $\left(V^{\perp}\right)^{\perp}=V$.

4. Confirm that the Fredholm alternative of this section holds for the system

$$\begin{aligned} x_1+x_2+2x_3&=5\\ 2x_1+3x_2-x_3&=2\\ 4x_1+5x_2+3x_3&=1 \end{aligned}$$

5. Use the subspace test to prove that if V is a subspace of the inner product space W, then so is $V^{\perp}$.

6. Show that if V is a subspace of the inner product space W, then $V\cap V^{\perp}=\{\mathbf{0}\}$.

7. Let U and V be subspaces of the inner product space W.

(a) Prove that $(U\cap V)^{\perp}=U^{\perp}+V^{\perp}$.

(b) Prove that $(U+V)^{\perp}=U^{\perp}\cap V^{\perp}$.

8. Find a QR factorization for the matrix $A=\begin{bmatrix} 1 & 2\\ -1 & 1\end{bmatrix}$.

9. Let

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

(a) Use the Gram-Schmidt algorithm to find a QR factorization of A.

(b) Use the result of (a) to find the least squares solution to the system $A\mathbf{x}=\mathbf{b}$, where $\mathbf{b}=(1,2,3)$.

10. Carry out the method of computing $U\cap V$ discussed on page 288 using these two subspaces of $W=\mathbb{R}^3$:

$$U=\operatorname{span}\{(1,2,1),(2,1,0)\}\qquad V=\operatorname{span}\{(1,1,1),(1,1,3)\}$$

11. The following is a simplified description of the *QR algorithm* (which is separate from the QR factorization, but involves it) for a real $n\times n$ matrix A :

$T_0=A,\ Q_0=I_n$

for $k=0,1,\dots$

$T_k=Q_{k+1}R_{k+1}$ (QR factorization of T_k)

$T_{k+1}=R_{k+1}Q_{k+1}$

end

Apply this algorithm to the following two matrices and, based on your results, speculate about what it is supposed to compute. You will need a CAS or MAS for this exercise and, of course, you will stop in a finite number of steps, but expect to take more than a few.

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & -2 \\ 0 & -2 & 1 \end{bmatrix} \qquad A = \begin{bmatrix} -8 & -5 & 8 \\ 6 & 3 & -8 \\ -3 & 1 & 9 \end{bmatrix}$$

6.5. *Operator Norms

The object of this section is to develop a useful notion of the norm of a matrix. For simplicity, we stick with the case of a real matrix A, but all of the results in this section carry over easily to complex matrices. In Chapters 3 and 5 we studied the concept of a vector norm, which gave us a way of thinking about the "size" of a vector. We could easily extend this to matrices, just by thinking of a matrix as a vector which had been chopped into segments of equal length and restacked as a matrix. Thus, every vector norm on the space $\mathbb{R}^{mn}$ of vectors of length mn gives rise to a vector norm on the space $\mathbb{R}^{m,n}$ of $m \times n$ matrices. Experience has shown that, with one exception – the standard norm, this is not the best way to look for norms of matrices. After all, matrices are deeply tied up with the operation of matrix multiplication. It would be too much to expect norms to distribute over products. The following definition takes a middle ground that has proved to be useful for many applications.

Matrix Norm

DEFINITION 6.5.1. A vector norm $||\cdot||$ which is defined on the vector space $\mathbb{R}^{m,n}$ of $m \times n$ matrices, for any pair m, n, is said to be a *matrix norm* if, for all pairs of matrices A, B which are conformable for multiplication,

$$||AB|| \le ||A|| \, ||B||$$

Our first example of such a norm is called the Frobenius norm; it is the one exception that we alluded to above.

DEFINITION 6.5.2. The *Frobenius norm* of a matrix $A = [a_{ij}]_{m,n}$ is defined by

$$||A||_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$$

THEOREM 6.5.3. *The Frobenius norm is a matrix norm.*

PROOF. Let A and B be matrices conformable for multiplication and suppose that the rows of A are $\mathbf{a}_1^T, \mathbf{a}_2^T, \ldots, \mathbf{a}_m^T$, while the columns of B are $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$. Then we have that $AB = [\mathbf{a}_i^T \mathbf{b}_j]$, so that by applying the definition and the CBS inequality, we obtain that

$$\begin{aligned} ||AB||_F = \left(\sum_{i=1}^m \sum_{j=1}^n \left| \mathbf{a}_i^T \mathbf{b}_j \right|^2 \right)^{1/2} &\leq \left(\sum_{i=1}^m \sum_{j=1}^n \|\mathbf{a}_i\|^2 \|\mathbf{b}_j\|^2 \right)^{1/2} \\ &\leq \left(\|A\|_F^2 \|B\|_F^2 \right)^{1/2} = \|A\|_F \|B\|_F \end{aligned}$$

□

The most important multiplicative norm comes from a rather general notion. Just as every inner product "induces" a norm in a natural way, every norm on the standard spaces induces a norm on matrices in a natural way. This type of norm is defined as follows.

DEFINITION 6.5.4. The *operator norm* induced on matrices by a norm on the standard spaces is defined by the formula

$$||A|| = \sup_{\mathbf{x} \neq 0} \frac{||A\mathbf{x}||}{||\mathbf{x}||}$$

A useful fact about these norms is the following equivalent form.

$$||A|| = \sup_{\mathbf{x} \neq 0} \frac{||A\mathbf{x}||}{||\mathbf{x}||} = \sup_{\mathbf{x} \neq 0} \left|\left| A \frac{\mathbf{x}}{||\mathbf{x}||} \right|\right| = \sup_{||\mathbf{v}||=1} ||A\mathbf{v}||$$

THEOREM 6.5.5. *Every operator norm is a matrix norm.*

PROOF. For a given matrix A clearly $||A|| \geq 0$ with equality if and only if $A\mathbf{x} = 0$ for all vectors x, which is equivalent to $A = 0$. The remaining two norm properties are left as exercises. Finally, if A and B are conformable for multiplication, then

$$||AB|| = \sup_{\mathbf{x} \neq 0} \frac{||AB\mathbf{x}||}{||\mathbf{x}||} \leq ||A|| \sup_{\mathbf{x} \neq 0} \frac{||B\mathbf{x}||}{||\mathbf{x}||} = ||A|| \cdot ||B||$$

□

Incidentally, one difference between the Frobenius norm and operator norms is how the identity I_n is handled. Notice that $\|I_n\|_F = n$, while with any operator norm $\|\cdot\|$, we have from the definition that $\|I_n\| = 1$.

How do we compute these norms? The next result covers the most common cases.

THEOREM 6.5.6. *Let* $A = [a_{ij}]_{m,n}$. *Then*

1. $||A||_1 = \max_{1 \leq i \leq m} \{\sum_{j=1}^n |a_{ij}|\}$
2. $||A||_\infty = \max_{1 \leq j \leq n} \{\sum_{i=1}^m |a_{ij}|\}$
3. $||A||_2 = \rho(A^T A)^{1/2}$

PROOF. Items (1) and (3) are left as an exercises. For the proof of (2), use the fact that $||A||_\infty = \sup_{||\mathbf{v}||_\infty = 1} ||A\mathbf{v}||_\infty$. Now a vector has infinity norm 1 if each of its coordinates is 1 in absolute value. Notice that we can make the ith entry of $A\mathbf{v}$ as large as possible simply by choosing $\mathbf{v}$ so that the jth coordinate of $\mathbf{v}$ is ± 1 and agrees with the sign of a_{ij}. Hence the infinity norm of $A\mathbf{v}$ is the maximum of the row sums of the absolute values of the entries of A, as stated in (2). □

One of the more important applications of the idea of a matrix norm is the famous Banach Lemma. Essentially, it amounts to a matrix version of the familiar geometric series encountered in calculus.

Banach Lemma

THEOREM 6.5.7. *Let M be a square matrix such that $\|M\| < 1$ for some operator norm $\|\cdot\|$. Then the matrix $I - M$ is invertible. Moreover,*

$$(I - M)^{-1} = I + M + M^2 + \cdots + M^k + \cdots$$

and $\left\|(I - M)^{-1}\right\| \leq 1/(1 - \|M\|)$.

PROOF. Form the familiar telescoping series

$$(I - M)\left(I + M + M^2 + \cdots + M^k\right) = I - M^{k+1}$$

so that

$$I - (I - M)\left(I + M + M^2 + \cdots + M^k\right) = M^{k+1}$$

Now by the multiplicative property of matrix norms and fact that $\|M\| < 1$

$$\left\|M^{k+1}\right\| \leq \|M\|^{k+1} \underset{k \to \infty}{\to} 0.$$

It follows that the matrix $\lim_{k\to\infty}\left(I + M + M^2 + \cdots + M^k\right) = N$ exists and that $I - (I - M)B = 0$, from which it follows that $B = (I - M)^{-1}$. Finally, note that

$$\begin{aligned}
\left\|I + M + M^2 + \cdots + M^k\right\| &\leq \|I\| + \|M\| + \|M\|^2 + \cdots + \|M\|^k \\
&\leq 1 + \|M\| + \|M\|^2 + \cdots + \|M\|^k \\
&\leq \frac{1}{1 - \|M\|}.
\end{aligned}$$

Now take the limit as $k \to \infty$ to obtain the desired result. □

A fundamental idea in numerical linear algebra is the notion of the *condition number* of a matrix A. Roughly speaking, the condition number measures the degree to which changes in A lead to changes in solutions of systems $A\mathbf{x} = \mathbf{b}$. A large condition number means that small changes in A may lead to large changes in $\mathbf{x}$. In the case of an invertible matrix A, the condition number of A is defined to be

$$\operatorname{cond}(A) = ||A||\ \left|\left|A^{-1}\right|\right|$$

Of course this quantity is norm dependent. In the case of an operator norm, the Banach lemma has a nice application.

COROLLARY 6.5.8. *If $A = I + N$, where $\|N\| < 1$, then*

$$\operatorname{cond}(A) \leq \frac{1 + \|N\|}{1 - \|N\|}$$

We leave the proof as an exercise.

We conclude with a very fundamental result for numerical linear algebra. The context is a more general formulation of the problem which is discussed in Section 6.6. Here is the scenario: suppose that we desire to solve the linear system $A\mathbf{x} = \mathbf{b}$, where A is invertible. Due to arithmetic error or possibly input data error, we end up with a value $\mathbf{x} + \delta\mathbf{x}$ which solves exactly a "nearby" system $(A + \delta A)(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}$. (It can be shown by using an idea called "backward error analysis" that this is really what happens when many algorithms are used to solve a linear system.) The question is, what is the size of the relative error $\|\delta x\| / \|x\|$? As long as the perturbation matrix $\|\delta A\|$ is reasonably small, there is a very elegant answer.

THEOREM 6.5.9. *Suppose that A is invertible, $A\mathbf{x} = \mathbf{b}$, $(A + \delta A)(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}$ and $\|A^{-1}\delta A\| = c < 1$ with respect to some operator norm. Then $A + \delta A$ is invertible and* Perturbation Theorem

$$\frac{|\ \delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\operatorname{cond}(A)}{1 - c}\left[\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}\right]$$

PROOF. That the matrix $I + A^{-1}\delta A$ follows from hypothesis and the Banach lemma. Expand the perturbed equation to obtain

$$(A + \delta A)(\mathbf{x} + \delta\mathbf{x}) = A\mathbf{x} + \delta A\mathbf{x} + A\delta\mathbf{x} + \delta A\,\delta\mathbf{x} = \mathbf{b} + \delta\mathbf{b}$$

Now subtract the terms $A\mathbf{x} = \mathbf{b}$ from each side solve for $\delta\mathbf{x}$ to obtain

$$(A + \delta A)\delta\mathbf{x} = A^{-1}(I + \delta A^{-1}A)\delta\mathbf{x} = -\delta A\mathbf{x} + \delta\mathbf{b}$$

so that

$$\delta\mathbf{x} = (I + \delta A^{-1}A)^{-1}A^{-1}\left[-\delta A\mathbf{x} + \delta\mathbf{b}\right]$$

Now take norms and use the additive and multiplicative properties and the Banach lemma to obtain

$$\|\delta\mathbf{x}\| \leq \frac{\|A^{-1}\|}{1 - c}\left[\|\delta A\mathbf{x}\| + \|\delta\mathbf{b}\|\right].$$

Next divide both sides by $\|x\|$ to obtain

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|A^{-1}\|}{1 - c}\left[\|\delta A\| + \frac{\|\delta\mathbf{b}\|}{\|\mathbf{x}\|}\right]$$

Finally, notice that $\|\mathbf{b}\| \leq \|A\|\|x\|$. Therefore, $1/\|\mathbf{x}\| \leq \|A\|/\|\mathbf{b}\|$. Replace $1/\|\mathbf{x}\|$ in the right hand side by $\|A\|/\|\mathbf{b}\|$ and factor out $\|A\|$ to obtain

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|A^{-1}\|\|A\|}{1 - c}\left[\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}\right]$$

which completes the proof, since by definition, $\operatorname{cond} A = \|A^{-1}\|\|A\|$. □

If we believe that the inequality in the perturbation theorem can be sharp (it can!), then it becomes clear how the condition number of the matrix A is a direct factor in how relative error in the solution vector is amplified by perturbations in the coefficient matrix.

6.5 Exercises

1. Let

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

Compute the Frobenius, $1-$, $2-$, and ∞-norms of A.

2. With A as in Exercise 1, compute the condition number of A using the infinity norm.

3. Prove Corollary 6.5.8 by making use of the Triangle inequality and the Banach lemma.

4. Use the Banach lemma to show that if A is invertible, then so is $A + \delta A$ provided that $\|A^{-1}\delta A\| < 1$.

5. Prove that for a square matrix A, $||A||_1 = \max_{1\leq i\leq m}\{\sum_{j=1}^{n} |a_{ij}|\}$.

6. Prove that for a square matrix A, $||A||_2 = \rho(A^T A)^{1/2}$

7. Example 6.6.4 gives an upper bound on the error propagated to the solution of a system due to right hand side error. How pessimistic is it? Experiment with several different erroneous right hand sides of your own choosing and compare the actual error with estimated error.

8. Let

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

(a) Use Householder matrices to find a full QR factorization of A.

(b) Use the result of (a) to find the least squares solution to the system $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = (1, 2, 3)$.

6.6. *Computational Notes and Projects

Error and Limit Measurements

We are going to consider a situation where infinity norms are both more natural to a problem and easier to use than the standard norm. This material is a simplified treatment of some of the concepts introduced in Section 6.5 and is independent of that section. The theorem below provides a solution to this question: how large an error in the solution to a linear system can there be, given that we have introduced an error in the right hand side whose size we can estimate? (Such an error might be due to experimental error or input error.) The theorem requires an extension of the idea of vector infinity norm to matrices for its statement.

DEFINITION 6.6.1. Let A be an $n \times n$ matrix whose rows are $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n$. The *infinity norm of the matrix* A is defined as

$$||A||_\infty = \max\{||\mathbf{r}_1||_1, ||\mathbf{r}_2||_1, \ldots, ||\mathbf{r}_n||_1\}$$

If, moreover, A is invertible, then the condition number of A is defined to be

$$\operatorname{cond}(A) = ||A||_\infty \left|\left|A^{-1}\right|\right|_\infty$$

EXAMPLE 6.6.2. Let $A = \begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix}$. Find $||A||_\infty$, $\left|\left|A^{-1}\right|\right|_\infty$ and $\operatorname{cond}(A)$.

SOLUTION. Here we see that $A^{-1} = \begin{bmatrix} 2 & -1 \\ -1/2 & 1/2 \end{bmatrix}$. From the preceding definition we obtain that

$$||A||_\infty = \max\{|1| + |2|, |1| + |4|\} = 5$$

and

$$\left|\left|A^{-1}\right|\right|_\infty = \max\left\{|2| + |-1|, \left|\frac{-1}{2}\right| + \left|\frac{1}{2}\right|\right\} = 3$$

so it follows that

$$\operatorname{cond}(A) = 5 \cdot 3 = 15.$$

□

THEOREM 6.6.3. *Suppose that the* $n \times n$ *matrix* A *is nonsingular,* $A\mathbf{x} = \mathbf{b}$ *and* $A(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}$. *Then*

$$\frac{||\delta\mathbf{x}||_\infty}{||\mathbf{x}||_\infty} \le \operatorname{cond}(A)\frac{||\delta\mathbf{b}||_\infty}{||\mathbf{b}||_\infty}$$

PROOF. Subtract the first equation of the statement of the theorem from the second one to obtain that

$$A\tilde{\mathbf{x}} - A\mathbf{x} = A(\tilde{\mathbf{x}} - \mathbf{x}) = \tilde{\mathbf{b}} - \mathbf{b} = \delta\mathbf{b}$$

from which it follows that

$$\delta\mathbf{x} = \tilde{\mathbf{x}} - \mathbf{x} = A^{-1}\delta\mathbf{b}$$

Now write $A^{-1} = [c_{ij}]$, $\delta\mathbf{b} = [d_i]$ and compute the ith coordinate of $\delta\mathbf{x}$:

$$(\delta\mathbf{x})_i = \sum_{j=1}^{n} c_{ij} d_j$$

so that if $\mathbf{r}_i = (c_{i1}, c_{i2}, \ldots, c_{in})$ is the ith row of A^{-1}, then

$$\begin{aligned}
|(\delta\mathbf{x})_i| &\le \sum_{j=1}^{n} |c_{ij}|\,|d_j| \\
&\le \max\{|d_1|, \ldots, |d_n|\} \sum_{j=1}^{n} |c_{ij}| \\
&\le \|\delta\mathbf{b}\|_\infty \|\mathbf{r}_i\|_1
\end{aligned}$$

Therefore,

$$\|\delta\mathbf{x}\|_\infty \le \|\delta\mathbf{b}\|_\infty \left\|A^{-1}\right\|_\infty \tag{6.6.1}$$

A similar calculation shows us that since $\mathbf{b} = A\mathbf{x}$,

$$\|\mathbf{b}\|_\infty \le \|\mathbf{x}\|_\infty \|A\|_\infty$$

Divide both sides by $\|\mathbf{b}\|_\infty \|\mathbf{x}\|_\infty$ and obtain that

$$\frac{1}{\|\mathbf{x}\|_\infty} \le \|A\|_\infty \frac{1}{\|\mathbf{b}\|_\infty} \tag{6.6.2}$$

Now multiply the inequalities 6.6.1 and 6.6.2 together to obtain the asserted inequality of the theorem. □

EXAMPLE 6.6.4. Suppose we wish to solve the nonsingular system $A\mathbf{x} = \mathbf{b}$ exactly, where the coefficient matrix A is as in Example 6.6.2 but the right hand side vector $\mathbf{b}$ is determined from measured data. Suppose also that the error of measurement is such that the ratio of the largest error in any coordinate of $\mathbf{b}$ to the largest coordinate of $\mathbf{b}$ (this ratio is called the *relative error*) is no more than 0.01 in absolute value. Estimate the size of the relative error in the solution.

SOLUTION. In matrix notation, we can phrase the problem in this manner: let the correct value of the right hand side be $\mathbf{b}$ and the measured value of the right hand side be $\tilde{\mathbf{b}}$, so that the error of measurement is the vector $\delta\mathbf{b} = \tilde{\mathbf{b}} - \mathbf{b}$. Rather than solving the system $A\mathbf{x} = \mathbf{b}$, we end up solving the system $A\tilde{\mathbf{x}} = \tilde{\mathbf{b}} = \mathbf{b} + \delta\mathbf{b}$, where $\tilde{\mathbf{x}} = \mathbf{x} + \delta\mathbf{x}$. The relative error in data is the quantity $\|\delta\mathbf{b}\|_\infty / \|\mathbf{b}\|_\infty$, while the relative error in the computed solution is $\|\delta\mathbf{x}\|_\infty / \|\mathbf{x}\|_\infty$. This sets up very nicely for an application of Theorem 6.6.3. Furthermore, we already calculated $\operatorname{cond}(A) = 15$ in Example 6.6.2. It follows that the relative error in the solution satisfies the inequality

$$\frac{\|\delta x\|_\infty}{\|x\|_\infty} \le 15 \cdot 0.01 = 0.15$$

In other words, the relative error in our computed solution could be as large as 15%. □

A Practical QR algorithm

In the preceding section we saw that the QR factorization can be used to solve systems including least squares. We also saw the factorization as a consequence of the Gram-Schmidt algorithm. As a matter of fact, the *classical* Gram-Schmidt algorithm which we have presented has certain numerical stability problems when used in practice. There is a so-called *modified* Gram-Schmidt algorithm that performs better. However, there is another approach to QR factorization that avoids Gram-Schmidt altogether. This approach uses the Householder matrices we introduced in Section 4.3. It is more efficient and stable than Gram-Schmidt. If you use a MAS to find the QR factorization of a matrix, it is likely that this is the method used by the system.

The basic idea behind this Householder QR is to use a succession of Householder matrices to zero out the lower triangle of a matrix, one column at a time. The key fact about Householder matrices is the following application of these matrices:

THEOREM 6.6.5. *Let $\mathbf{x}, \mathbf{y}$ be nonzero vectors in $\mathbb{R}^n$ of the same length. Then there is a Householder matrix $H_{\mathbf{v}}$ such that $H_{\mathbf{v}}\mathbf{x} = \mathbf{y}$.*

PROOF. Let $\mathbf{v} = \mathbf{x} - \mathbf{y}$. Then we see that

$$(\mathbf{x}+\mathbf{y})^T(\mathbf{x}-\mathbf{y}) = \mathbf{x}^T\mathbf{x} - \mathbf{x}^T\mathbf{y} - \mathbf{y}^T\mathbf{x} + \mathbf{y}^T\mathbf{y} = \mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{y} = 0$$

since $\mathbf{x}$ and $\mathbf{y}$ have the same length. Now write

$$x = \frac{1}{2}\{(\mathbf{x}-\mathbf{y}) + (\mathbf{x}+\mathbf{y})\} = \mathbf{p} + \mathbf{u}$$

and obtain from Theorem 4.3.10 that

$$H_{\mathbf{v}}\mathbf{x} = -\mathbf{p} + \mathbf{u} = \frac{1}{2}\{-(\mathbf{x}-\mathbf{y}) + (\mathbf{x}+\mathbf{y})\} = \frac{2\mathbf{y}}{2} = \mathbf{y}$$

which is what we wanted to show. □

Now we have a tool for massively zeroing out entries in a vector of the form $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Set $y = (\pm\|\mathbf{x}\|, 0, \ldots, 0)$ and apply the preceding theorem to construct Householder H such that $H_{\mathbf{v}}\mathbf{x} = \mathbf{y}$. It is standard to choose the $\pm$ to be the negative of the sign of x_1. In this way, the first term will not cause any loss of accuracy to subtractive cancellation. However, any choice of $\pm$ works fine in theory. We can picture this situation schematically very nicely by representing possibly nonzero entries by an '$\times$' in the following simple version:

$$\mathbf{x} = \begin{bmatrix} \times \\ \times \\ \times \\ \times \end{bmatrix} \xrightarrow[H_{\mathbf{v}}]{} \begin{bmatrix} \pm\|\mathbf{x}\| \\ 0 \\ 0 \\ 0 \end{bmatrix} = H_{\mathbf{v}}\mathbf{x}$$

We can extend this idea to zeroing out lower parts of $\mathbf{x}$ only, say

$$\mathbf{x} = \begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{z} \\ \times \\ \times \\ \times \end{bmatrix} \text{by using} \mathbf{y} = \begin{bmatrix} \mathbf{z} \\ \pm\|\mathbf{w}\| \\ 0 \\ 0 \end{bmatrix} \text{so} \mathbf{v} = \begin{bmatrix} 0 \\ \times \\ \times \\ \times \end{bmatrix} \text{and} H_{\mathbf{v}}\mathbf{x} = \begin{bmatrix} 0 \\ \times \\ 0 \\ 0 \end{bmatrix}.$$

We can apply this idea to systematically zero out subdiagonal entries by successive multiplication by Householder (hence orthogonal) matrices; schematically we have this representation of a full rank $m \times n$ matrix A

$$A = \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \xrightarrow[H_1]{} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix} \xrightarrow[H_2]{} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & \times \end{bmatrix}$$

$$\xrightarrow[H_3]{} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & 0 \end{bmatrix} = R$$

so that $H_3H_2H_1A = R$. Now we can check easily from the definition of a Householder matrix H that $H^T = H = H^{-1}$. Thus, if we set $Q = H_1^{-1}H_2^{-1}H_3^{-1} = H_1H_2H_3$, it follows that $A = QR$. Notice that we don't actually have to carry out the multiplications to compute Q unless they are needed, and the vectors needed to define these Householder matrices are themselves easily stored in a single matrix. What we have here is just a bit different from the QR factorization discussed in the last section. Here the matrix Q is a full $m \times m$ matrix and R is the same size as A. Even if A is not full column rank, this procedure will work, provided we simply skip construction of H in the case that there are no nonzero elements to zero out in some column. Consequently, we have essentially proved the following theorem, which is sometimes called a *full* QR factorization, in contrast to the *reduced* QR factorization of Theorem 6.4.10.

Full QR Factorization

THEOREM 6.6.6. *Let A be a real $m \times n$ matrix. Then there exists an $m \times m$ orthogonal matrix Q and $m \times n$ upper triangular matrix R such that $A = QR$.*

Actually, all of the results we have discussed regarding QR factorization carry over to complex matrices, provided we replace orthogonal matrices by unitary matrices and transposes by Hermitian transposes.

Project Topics

Project: Testing Least Squares Solvers

The object of this project is to test the quality of the solutions of three different methods for solving least squares problems $A\mathbf{x} = \mathbf{b}$:

(a) Solution by solving the associated normal equations by Gaussian elimination.

(b) Solution by reduced QR factorization obtained by Gram-Schmidt.

(c) Solution by full QR factorization by Householder matrices.

Here is the test problem: suppose we want to approximate the curve $f(x) = e^{\sin(6x)},\ 0 \leq x \leq 1$ by a tenth degree polynomial. The input data will be the sampled values of $f(x)$ at equally spaced nodes $x = kh,\ k = 0, 1, \ldots, 20,\ h = 0.05$. The fact that

$$f(x) = c_0 + c_1x + \cdots c_{10}x^{10}$$

gives 21 equations for the 11 unknown coefficients $c_k,\ k = 0, 1, \ldots, 20$. The coefficient matrix that results from this problem is called a Vandermonde matrix. Your MAS should have a have a built-in command for construction of such a matrix.

Procedure: First set up the system matrix A and right hand side matrix b. Method (a) is easily implemented on any CAS or MAS. The built-in procedure for computing a QR factorization will very likely be Householder matrices which will take care of (c). You will need to check the documentation to verify this. The Gram-Schmidt method of finding QR factorization will have to be programmed by you.

Once you have solved the system by these three methods, make out a table that has the computed coefficients for each of the three methods. Then make plots of the difference between the function $f(x)$ and the computed polynomial for each method. Discuss your results.

There are a number of good texts which discuss numerical methods for least squares; see, e.g., [**3**]

Project: Approximation Theory

Suppose you work for a manufacturer of calculators, and are involved in the design of a new calculator. The problem is this : as one of the "features" of this calculator, the designers decided that it would be nice to have a key which calculated a transcendental function, namely, $f(x) = \sin(\pi x), -1 \leq x \leq 1$ Your job is to come up with an adequate way of calculating $f(x)$, say with an error no worse than $.001$

Polynomials are a natural idea for approximating functions. From a designer's point of view they are particularly attractive because they are so easy to implement. Given the coefficients of a polynomial, it is easy to design a very efficient and compact algorithm for calculating values of the polynomial. Such an algorithm, together with the coefficients of the polynomial, would fit nicely into a small ROM for the calculator, or could even be microcoded into the chip.

Your task is to find a low degree polynomial that approximates $\sin(\pi x)$ to within the specified accuracy. For comparison, find a Taylor polynomial of lowest degree for $\sin x$ that gives sufficient accuracy. Next, use the projection problem idea to project the function $\sin x \in C[-1,1]$ with the standard inner product, into the subspace $\mathcal{P}_n$ of polynomials of degree at most n. You will need to find the smallest n that gives a projection whose difference from $\sin x$ is at most 0.001 on the interval $[-1,1]$. Is it lower degree than the best Taylor polynomial approximation?

Use a CAS to do the computations and graphics. Then report on your findings. Include graphs that will be helpful in interpreting your conclusions. Also, give suggestions on how to compute this polynomial efficiently.

A Report Topic: Fourier Analysis

This project will introduce you to a very fascinating and important topic known as Fourier analysis. The setting is as follows: we are interested in finding approximations to functions in the vector space $\mathcal{C}_{2\pi}$ of continuous periodic functions on the closed interval $[-\pi,\pi]$. This vector space becomes an inner product space with the usual definition

$$\langle f,g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,dx.$$

In this space the sequence of trigonometric functions

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \dots, \frac{\cos kx}{\sqrt{\pi}}, \frac{\sin kx}{\sqrt{\pi}}, \dots$$

forms an orthonormal set. Therefore, we can form the finite dimensional subspaces V_n spanned by the first $2n+1$ of these elements and immediately obtain an orthonormal basis of V_n. We can also use the machinery of projections to approximate any function $f(x) \in \mathcal{C}_{2\pi}$ by its projection into the various subspaces V_n. The coefficients of the orthonormal basis functions in the projection formula of Definition 6.3.5 as applied to a function $f(x)$ are called the *Fourier coefficients* of $f(x)$. They are traditionally designated by the symbols

$$\frac{a_0}{2}, a_1, b_1, a_2, b_2, \ldots, a_k, b_k, \ldots$$

In the first part of this project you will write a brief introduction to Fourier analysis in which you exhibit formulas for the Fourier coefficients of a function $f(x)$ and explain the form and meaning of the projection formula in this setting. Try to prove that the trigonometric functions given above are an orthonormal set. At minimum provide a proof for the first three functions.

In the second part you will explore the quality of these approximations for various test functions. The test functions are specified on the interval $[-\pi, \pi]$ and then this graph is replicated on adjacent intervals of length 2π, so they are periodic.

1. $f(x) = \sin \frac{x^2}{\pi}$
2. $g(x) = x(x-\pi)(x+\pi)$
3. $h(x) = x$

Notice that the last function violates the continuity condition.

For each test function you should prepare a graph that includes the test function and at least two projections of it into the V_n, $n = 0, 1, \ldots$. Discuss the quality of the approximations and report on any conclusions that you can draw from this data. You will need a MAS or CAS to carry out the calculations and graphs, as the calculations are very detailed. If you are allowed to do so, you could write your report up in the form of a notebook.

6.6 Exercises

1. Example 6.6.4 gives an upper bound on the error propagated to the solution of a system due to right hand side error. How pessimistic is it? Experiment with several different erroneous right hand sides of your own choosing and compare the actual error with estimated error.

2. Let

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

(a) Use Householder matrices to find a full QR factorization of A.

(b) Use the result of (a) to find the least squares solution to the system $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = (1, 2, 3)$.

Review

Chapter 6 Exercises

1. Let the vector space $V = \mathbb{R}^4$, equipped with the standard inner product.

(a) Apply the Gram-Schmidt algorithm to vectors $(1,1,1,1), (4,2,4,2), (0,0,0,2)$ to obtain an orthogonal set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

(b) Normalize the orthogonal list obtained to obtain an orthonormal set $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$.

(c) Use this to find the projection matrix for the subspace spanned by these vectors.

2. Let

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

(a) Use Householder matrices to find a full QR factorization of A.

(b) Use the result of (a) to find the least squares solution to the system $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = (1,2,3)$.

3. Determine if the formula $\langle (u,v),\ (x,y) \rangle = ux - vy$ defines an inner product on $\mathbb{R}^2$.

4. Find the projection matrix into the row space of the matrix $\begin{bmatrix} 3 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$.

5. Use the Gram-Schmidt algorithm to expand the orthogonal set $(1,1,1)$, $(1,-2,1)$ into an orthogonal basis of $\mathbb{R}^3$ and then normalize this basis to obtain an orthonormal basis.

6. Suppose that A is an $n \times n$ matrix of rank $n-1$. Show that all rows of $\operatorname{adj} A$ are multiples of each other. *Hint:* Use the adjoint formula and orthogonal complements.

7. Find the orthogonal complement of $V = \operatorname{span}\{1 + x + x^2\}$ in $W = \mathcal{P}_2$ where W has the usual function space inner product $f, g = \int_0^1 f(x)g(x)\,dx$.

8. Given an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ of $\mathbb{R}^3$, show that

$$\mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T + \mathbf{u}_3\mathbf{u}_3^T = I$$

where each term is a projection matrix in the sense of Exercise 8 and the product of any two distinct terms is 0. (Such an expression is called a *resolution of the identity* .)

APPENDIX A

Table of Symbols

Symbol	Meaning	Reference
$\emptyset$	Empty set	Page 9
$\in$	Member symbol	Page 9
$\subseteq$	Subset symbol	Page 9
$\subset$	Proper subset symbol	Page 9
$\cap$	Intersection symbol	Page 9
$\cup$	Union symbol	Page 9
$\overrightarrow{PQ}$	Displacement vector	Page 127
$\lvert z \rvert$	Absolute value of complex z	Page 12
$\lvert A \rvert$	determinant of matrix A	Page 98
$\lvert\lvert \mathbf{u} \rvert\rvert$	Norm of vector $\mathbf{u}$	Page 185
$\lvert\lvert \mathbf{u} \rvert\rvert_p$	p-norm of vector $\mathbf{u}$	Page 262
$\mathbf{u} \cdot \mathbf{v}$	Standard inner product	Page 189
$\langle \mathbf{u}, \mathbf{v} \rangle$	Inner product	Page 267
$\operatorname{adj} A$	Adjoint of matrix A	Page 106
A^H	Hermitian transpose of matrix A	Page 77
A^T	Transpose of matrix A	Page 77
$\mathcal{C}(A)$	Column space of matrix A	Page 153
$\operatorname{cond}(A)$	Condition number of matrix A	Page 297
$C[a, b]$	Function space	Page 129
$\mathbb{C}$	Complex numbers $a + bi$	Page 11
$\mathbb{C}^n$	Standard complex vector space	Page 127
$\operatorname{comp}_{\mathbf{v}} \mathbf{u}$	Component	Page 196
A_{cof}	Cofactor matrix of A	Page 11
$\overline{z}$	Complex conjugate of z	Page 12
δ_{ij}	Kronecker delta	Page 106
$\dim V$	Dimension of space V	Page 163
$\det A$	Determinant of A	Page 97
$\operatorname{domain}(T)$	Domain of operator T	Page 158
$\operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$	Diagonal matrix with entries $\lambda_1, \lambda_2, \ldots, \lambda_n$ on diagonal	Page 225
E_{ij}	Elementary row operation switching ith and jth rows	Page 22
$E_i(c)$	Elementary row operation multiplying ith row by c	Page 22
$E_{ij}(d)$	Elementary operation adding d times jth row to ith row	Page 22

Symbol	Meaning	Reference
$\mathcal{E}_\lambda(A)$	Eigenspace	Page 216
$H_\mathbf{v}$	Householder matrix	Page 206
I, I_n	Identity matrix, $n \times n$ identity	Page 73
$\Im(z)$	Imaginary part of z	Page 11
$\ker(T)$	Kernel of operator T	Page 158
$M_{ij}(A)$	Minor of A	Page 99
$M(A)$	Matrix of minors of A	Page 106
$\max\{a_1, a_2, \ldots, a_m\}$	Maximum value	Page 35
$\min\{a_1, a_2, \ldots, a_m\}$	Minimum value	Page 35
$\mathcal{N}(A)$	Null space of matrix A	Page 153
$\mathbb{N}$	Natural numbers $1, 2, \ldots$	Page 10
$\mathrm{null}\, A$	Nullity of matrix A	Page 35
$\mathcal{P}$	Space of polynomials of any degree	Page 137
$\mathcal{P}_n$	Space of polynomials of degree $\leq n$	Page 137
$\mathrm{proj}_\mathbf{v}\, \mathbf{u}$	Projection vector along a vector	Page 196
$\mathrm{proj}_V\, \mathbf{u}$	Projection vector into subspace	Page 281
$\mathbb{Q}$	Rational numbers a/b	Page 10
$\Re(z)$	Real part of z	Page 11
$\mathcal{R}(A)$	Row space of matrix A	Page 153
$R(\theta)$	Rotation matrix	Page 149
$\mathbb{R}$	Real numbers	Page 10
$\mathbb{R}^n$	Standard real vector space	Page 127
$\mathbb{R}^{m,n}$	Space of $m \times n$ real matrices	Page 128
$\mathrm{range}(T)$	Range of operator T	Page 158
$\mathrm{rank}\, A$	Rank of matrix A	Page 34
$\rho(A)$	Spectral radius of A	Page 232
$\mathrm{span}\{S\}$	Span of vectors in S	Page 138
$\mathrm{target}(T)$	Target of operator T	Page 158
$V^\perp$	Orthogonal complement of V	Page 286
$\mathbb{Z}$	Integers $0, \pm 1, \pm 2, \ldots$	Page 10
$[T]_{B,C}$	Matrix of operator T	Page 175
$\otimes$	tensor symbol	Page 114

APPENDIX B

Solutions to Selected Exercises

Section 1.1, Page 7

1 (a) $x = -1$, $y = 1$. (b) $x = 2$, $y = -2$, $z = 1$. (c) $x = 2$, $y = 1$

2. (a) is linear, and (b) is not linear. (a) in standard format is $\begin{array}{r} x - y - z = -2 \\ 3x - y = 4 \end{array}$.

3. (a) $m = 3$, $n = 3$, $a_{11} = 1$, $a_{12} = -2$, $a_{13} = 1$, $b_1 = 3$, $a_{21} = 0$, $a_{22} = 1$, $a_{23} = 0$, $b_2 = 2$, $a_{31} = -1$, $a_{32} = 0$, $a_{33} = 1$, $b_3 = 1$.

4. $$\begin{aligned} 2y_1 - y_2 &= \tfrac{1}{49} f(1/7) \\ -y_1 + 2y_2 - y_3 &= \tfrac{1}{49} f(2/7) \\ -y_2 + 2y_3 - y_4 &= \tfrac{1}{49} f(3/7) \\ -y_3 + 2y_4 - y_5 &= \tfrac{1}{49} f(4/7) \\ -y_4 + 2y_5 - y_6 &= \tfrac{1}{49} f(5/7) \\ -y_5 + 2y_6 &= \tfrac{1}{49} f(6/7) \end{aligned}$$

5. The equations are $\begin{array}{r} .8x - .1z = 2 \\ -.4x + .9z = 3 \end{array}$ so $x = \frac{105}{34}$, $z = \frac{80}{17}$.

6. $$\begin{aligned} -.8x + .1y + .4z + .4w &= 0 \\ .3x - .6y + .2z + .1w &= 0 \\ .3x + .4y - .8z + .3w &= 0 \\ .2x + .1y + .2z - .8w &= 0 \end{aligned}$$

7. $$\begin{aligned} a + b + c &= 1 \\ a + 2b + 4c &= 1 \\ a + 3b + 9c &= 2 \end{aligned}$$

The equation that comes from vertex v_1 is $-x_1 + x_4 - x_5 = 0$, and vertex v_3 is $x_2 - x_3 = 0$.

Section 1.2, Page 17

1. (a) $\{0, 1\}$ (b) $\{x | x \in \mathbb{Z}$ and $x > 1\}$ (c) $\{x | x \in \mathbb{Z}$ and $x \le -1\}$ (d) B

2. (a) $e^{3\pi i/2}$, (b) $\sqrt{2} e^{\pi i/4}$, (c) $2e^{2\pi i/3}$, (d) $e^{\pi i}$, (e) $2\sqrt{2} e^{7\pi i/4}$

3. (a) $1 + 8i$, (b) $\frac{3}{5} + \frac{4}{5} i$, (c) $10 + 10i$, (d) $\frac{-3}{5} + \frac{4}{5} i$

4. (a) $\frac{2}{5} - \frac{i}{5}$, (b) $-2 + i$, (c) $\{a + bi | a, b \in \mathbb{R}$ and $b = 2a + 1\}$

5. (a) $z = \frac{-1}{2} \pm \frac{\sqrt{11}}{2}i$, (b) $z = \frac{\pm\sqrt{3}}{2} + \frac{i}{2}$, (c) $z = 1 \pm (\frac{-\sqrt{2\sqrt{2}+2}}{2} + \frac{\sqrt{2\sqrt{2}-2}}{2}i)$

6. (a) $z = e^0, e^{2\pi i/3}, e^{4\pi i/3}$, $z = 1, \frac{-1}{2} + \frac{\sqrt{3}}{2}i, \frac{-1}{2} - \frac{\sqrt{3}}{2}i$, (b) $z = 2e^{\pi i}, 2e^{\pi i/3}, 2e^{5\pi i/3}$, $z = -2, 1+\sqrt{3}i, 1-\sqrt{3}i$

11. Let $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$. Then $z_1z_2 = (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1)$. Thus

$$\begin{aligned} |z_1z_2|^2 &= (x_1x_2 - y_1y_2)^2 + (x_1y_2 + x_2y_1)^2 \\ &= x_1^2x_2^2 + y_1^2y_2^2 + x_1^2y_2^2 + x_2^2y_1^2 \\ &= (x_1^2 + y_1^2)(x_2^2 + y_2^2) \\ &= |z_1|^2|z_2|^2. \end{aligned}$$

Since $|z_1|$, $|z_2|$, and $|z_1z_2|$ all have positive values, it follows then that $|z_1z_2| = |z_1||z_2|$.

12.
$$\begin{aligned} e^{i(\theta+\psi)} &= \cos(\theta+\psi) - i\sin(\theta+\psi) \\ &= \cos\theta\cos\psi - \sin\theta\sin\psi \\ &\quad -i\sin\theta\cos\psi - i\cos\theta \\ &= \sin\psi \\ &= (\cos\theta - i\sin\theta)(\cos\psi - i\sin\psi) \\ &= e^{i\theta}e^{i\psi} \end{aligned}$$

13. The equation should have 5 roots.

Section 1.3, Page 27

2 (a) $x = 20, y = -11$, (b) $x_1 = 3, x_2 = -2, x_3 = 1$, (c) $x_1 = 3, x_2 = -5$.

3. Each augmented matrix is 3×5(a) $A = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 2 & 2 & 1 & 1 & 1 \\ 2 & 2 & 0 & 2 & 2 \end{bmatrix}$, size of A is 3×5, the general solution is $\begin{array}{l} x_1 = 1 - x_2 - x_4 \\ x_3 = x_4 - 1 \\ x_2, x_4 \text{ are free.} \end{array}$

(b) $A = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ -2 & -4 & 1 & 0 & 0 \\ 3 & 6 & -1 & 1 & 0 \end{bmatrix}$, size of A is 3×5, the general solution is
$x_1 = -2x_2$
$x_2 = 0$
$x_3 = 0$
x_2 is free.

4. (a) $\begin{array}{l} x_1 = 4 \\ x_3 = 2 \\ x_2 \text{ is free} \end{array}$ (b) $\begin{array}{l} x_1 = 1 \\ x_2 = 2 \\ x_3 = 2 \end{array}$ (c) is inconsistent.

5. (a) $\begin{array}{ll} x_1 = & \frac{2}{3}b_1 + \frac{1}{3}b_2 \\ x_2 = & \frac{-1}{3}b_1 + \frac{1}{3}b_2 \end{array}$ (b) If $b_2 - 2b_1 \neq 0$, then the system is inconsistent. Otherwise, the solution is $x_1 = b_1 + x_2$ and x_2 arbitrary.

(a)
$$\begin{array}{rcl} 2x_1 + x_2 + 7x_3 &=& -1 \\ 3x_1 + 2x_2 - 2x_4 &=& 1 \\ 2x_1 + 2x_2 + 2x_3 - 2x_4 &=& 4 \end{array}$$

(b)
$$\begin{array}{rcl} x_1 + x_2 + x_3 - x_4 &=& 2 \\ 2x_1 + x_2 - 2x_4 &=& 1 \\ 2x_1 + 2x_2 + 2x_3 - 2x_4 &=& 4 \end{array}$$

6. (b) $\begin{array}{l} x_1 = -1 + x_3 + x_4 \\ x_2 = 3 - 2x_3 \\ x_3, x_4 \text{ are free.} \end{array}$

7. $\begin{array}{l} x = \frac{131}{85}w \\ y = \frac{128}{85}w \\ z = \frac{29}{17}w \end{array}$ is the system's solution which has a nontrivial solution consisting of positive numbers if $w > 0$.

8. Applying the same operations to b, you find $b' = \begin{bmatrix} b_2 - b_1 \\ 2b_1 - b_2 \\ b_3 - 2b_1 \end{bmatrix}$. Since the bottom row in the coefficient matrix is all zeros after Gauss-Jordan elimination, the system is only consistent when $b_3 - 2b_1 = 0$.

9. (a) $\begin{array}{l} x_1 = 1 \\ x_2 = 0 \\ x_3 = 0 \end{array}$ (b) $\begin{array}{l} x_1 = 0 \\ x_2 = 0 \\ x_3 = 0 \end{array}$ (c) $\begin{array}{l} x_1 = 1 \\ x_2 = 1 \\ x_3 = 1 \end{array}$

(a) $\begin{array}{rcl} 2x - y + 3xy &=& 0 \\ 4x + 2y - xy &=& 0 \end{array}$ (b)
$$\begin{array}{rcl} yz + 3xz - xy &=& 0 \\ yz + 2xy &=& 0 \end{array}$$

10 (a) If $x = 0$ then the system reduces to $y = 0$. Similarly for $y = 0$. So $(0, 0)$ is a solution. If neither x nor y is zero, then divide both equations by xy to get a linear system in the variables $1/x, 1/y$, then solve this system.

11. $y_1 = \frac{40}{2187}$, $y_2 = \frac{77}{2187}$, $y_3 = \frac{4}{81}$, $y_4 = \frac{130}{2187}$, $y_5 = \frac{140}{2187}$, $y_6 = \frac{5}{81}$, $y_7 = \frac{112}{2187}$, $y_8 = \frac{68}{2187}$

15. Physically meaningful solutions are those that are nonnegative. Coefficient matrix of the system is
$$\begin{bmatrix} 1 & 0 & 0 & -1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix}.$$

Section 1.4, Page 38

1. (d) and (e) are in reduced row echelon form.

2. (a) $\begin{bmatrix} 1 & 0 & \frac{5}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}$, $\operatorname{rank} = 2$, $nullity = 1$.
(b) $\begin{bmatrix} 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & 0 & -33 \\ 0 & 0 & 1 & 2 \end{bmatrix}$, $\operatorname{rank} = 3$, $nullity = 1$.
(c) $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$, $\operatorname{rank} = 2$, $nullity = 2$.
(d) $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$, $\operatorname{rank} = 2$, $nullity = 1$.
(e) $\begin{bmatrix} 1 & 1 & 0 & \frac{22}{9} \\ 0 & 0 & 1 & \frac{2}{9} \end{bmatrix}$, $\operatorname{rank} = 2$, $nullity = 2$.

3. Let A be the augmented matrix and C be the coefficient matrix. (a) $\operatorname{rank}(A) = 2$, $\operatorname{rank}(C) = 2$, (b) $\operatorname{rank}(A) = 3$, $\operatorname{rank}(C) = 3$

5. (a) 2, (b) 0, (c) 3, (d) 1

6. $0 \le \operatorname{rank}(A) \le 2$

7. (a) true, (b) true, (c) false

8. $\begin{bmatrix} 1 & \frac{b}{a} \\ 0 & 0 \end{bmatrix}$

9 The condition is that the $m \times n$ coefficient matrix of the system have rank n.

Section 1.5, Page 47

2. (a) $\operatorname{rank}(A) = 3$, (b) $\begin{bmatrix} 1 & 3 & 0 & 4 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

3. The total work done at the nth stage is $n + 2\left[1 + 2 + \ldots + (n-1)\right] = \frac{n^2+n}{2}$.

Section 1.6, Page 47

1. (a) $2\sqrt{5}$, (b) $7 + 6i$

2. (a) $z = \frac{6}{5} - \frac{8}{5}i$, (b) $z = \pm\sqrt{2} \pm \sqrt{2}i$

3. (a) $\frac{\sqrt{2}}{2}e^{\pi i/4}$, $\frac{1}{2} + \frac{i}{2}$

5. Let A be the augmented matrix of the system.
(a) $\operatorname{rref}(A) = \begin{bmatrix} 1 & 0 & 0 & -4 & 2 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$.
The system's solution is
$$\begin{aligned} x_1 &= 2 + 4x_4 \\ x_2 &= -2x_4 \\ x_3 &= 1 - x_4 \\ x_4 &\text{ is free.} \end{aligned}$$
(b) $\operatorname{rank}(A) = 3$, $\operatorname{rref}(A) = E_{12}(-2)E_{23}(1)E_2(-1)E_3(\frac{1}{2})E_{21}(-1)A$.
(c) Since the rank of the coefficient matrix is 3, the coefficient matrix augmented with any right hand side will always be consistent and therefore have a solution.

6. (a) 3, (b) true, (c) true, (d) 1

Section 2.1, Page 54

2. (a) not possible, (b) $\begin{bmatrix} -1 & 0 & 1 \\ -1 & -1 & -2 \end{bmatrix}$,
(c) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

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

4 (a) $x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 2 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

8. Let $A = [a_{ij}]$, $B = [b_{ij}]$, $C = [c_{ij}]$. So

$$\begin{aligned}(A+B)+C &= [a_{ij} + b_{ij}] + [c_{ij}] \\ &= [a_{ij} + b_{ij} + c_{ij}] \\ &= [a_{ij}] + [b_{ij} + c_{ij}] \\ &= A + (B + C).\end{aligned}$$

9. Let $A = [a_{ij}]$, $B = [b_{ij}]$, and c, d be scalars. So

$$\begin{aligned}(c+d)A &= [(c+d)a_{ij}] \\ &= [ca_{ij} + da_{ij}] \\ &= [ca_{ij}] + [da_{ij}] \\ &= cA + dA.\end{aligned}$$

Similarly, you can show $c(A + B) = cA + cB$.

Section 2.2, Page 60

1. (a) $\begin{bmatrix} 1 & -2 & 4 \\ 0 & 1 & -1 \\ -1 & 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & -1 & -3 \\ 2 & 2 & 4 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 10 \\ 3 \end{bmatrix}$

2. $(I + A + A^2)(I - A) = \begin{bmatrix} -1 & -6 \\ -3 & -4 \end{bmatrix} = I - A^3$

3. (a) $x_1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 4 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$,

(b) $x \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + z \begin{bmatrix} -3 \\ 4 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ 10 \\ -3 \end{bmatrix}$

4. (a) $[19 + 3i]$, (b) $\begin{bmatrix} 6 & 8 \\ 3 & 4 \end{bmatrix}$, (c) impossible

5. (a) $\begin{bmatrix} 3 & -2 & 0 \\ 6 & -4 & 0 \end{bmatrix}$, (b) not possible, (c) $\begin{bmatrix} -15 & 15 & 3 \\ -5 & 4 & -2 \\ -24 & 25 & 3 \end{bmatrix}$

6. (a) $A = \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$ work.

7. Let $A_{m \times n} = [a_{ij}]$ and $B_{m \times n} = [b_{ij}]$. If $\mathbf{b} = [1, 0, \ldots, 0]^T$, $\begin{bmatrix} a_{11} & 0 \cdots 0 \\ \vdots & \vdots \\ a_{m1} & 0 \cdots 0 \end{bmatrix} = \begin{bmatrix} b_{11} & 0 \cdots 0 \\ \vdots & \vdots \\ b_{m1} & 0 \cdots 0 \end{bmatrix}$ so $a_{11} = b_{11}$, etc. By similar computations, you can show that for each ij, $a_{ij} = b_{ij}$ so $A = B$.

8. (b) is not nilpotent, the others are. For (d) $A^3 = 0$ so A is nilpotent.

12. Let $A = [a_{ij}]$, $B = [b_{ij}]$, and c be a scalar. Then the ij^{th} entry of $c(AB) = c(\sum_k a_{ik} b_{kj})$. The ij^{th} entry of $(cA)B = \sum_k (ca_{ik}) b_{kj} = c \sum_k a_{ik} b_{kj}$ = the ij^{th} entry of $c(AB)$. So

14. $f(A) = \begin{bmatrix} -12 & 10 \\ -5 & -12 \end{bmatrix}$.

Section 2.3, Page 71

6. (a) $[0.3, 0.4, 0.3]^T$, (b) the system does tend toward $\left[\frac{5}{29}, \frac{15}{29}, \frac{9}{29}\right]^T$

8. (2) $A = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$

(3)

vertex	power
1	2
2	4
3	3
4	5
5	3

Vertices 4 is the strongest.

9. (b)

node	power
1	7
2	8
3	10
4	11
5	10

, (b) 939

Section 2.4, Page 82

2. (a) Interchange columns 1 and 2 of A.

3. (a) upper triangular, triangular, strictly triangular and tridiagonal.

8. (a) $E_{12}(3)$, (b) $E_{ij}(-a)$, (c) $E_{31}(2)$, (d) $E_2(3)$

9. (a) $I_2 = E_{12}(-2)E_{21}(-1)\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$

10. (a) $\begin{bmatrix} 1 & -3 & 2 \end{bmatrix}^T = \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}$, so this matrix is not symmetric or Hermitian.

11. $(E_{23}(4))^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix} = E_{32}(4)$

12. (a) false, (b) true, (c) false, (d) false, (e) false

16. Let $Q(x, y, z) = \mathbf{x}^T A\mathbf{x}$ where $\mathbf{x} = [x, y, z]^T$ and $A = \begin{bmatrix} 2 & 2 & -6 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$.

18. $(A^H A)^H = A^H (A^H)^H = A^H A$ so $A^H A$ is Hermitian. Similarly, you can show AA^H is Hermitian.

19. Since two vectors, $\mathbf{a}$ and $\mathbf{b}$, will each have $\operatorname{rank} \leq 1$, $\operatorname{rank}(\mathbf{ab}^T) \leq \min\{\operatorname{rank}(\mathbf{a}), \operatorname{rank}(\mathbf{b}^T)\} \leq 1$.

23. If $N = \begin{bmatrix} 0 & & * \\ \vdots & \ddots & \\ 0 & \cdots & 0 \end{bmatrix}$, $N^2 = \begin{bmatrix} 0 & \mathbf{0} & & * \\ \vdots & \ddots & \ddots & \\ \vdots & & \ddots & \mathbf{0} \\ 0 & \cdots & \cdots & 0 \end{bmatrix}$ so N^2 has zeros along the first superdiagonal (in bold). By similar computation, N^3 has zeros along the first and second superdiagonals, etc. Therefore N^{n-1} has zeros along all of its superdiagonals so $N^{n-1} = \mathbf{0}$.

Section 2.5, Page 94

1. (a) $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ (b) $\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 1 & -1 & 1 \end{bmatrix}$ (c) inverse does not exist, (d) $\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 1 & -1 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ (e) $\begin{bmatrix} \frac{6}{13} + \frac{4}{13}i & \frac{-4}{13} - \frac{7}{13}i \\ \frac{2}{13} - \frac{3}{13}i & \frac{3}{13} + \frac{2}{13}i \end{bmatrix}$

2. (a) $A^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$.

3. (a) $\mathbf{x} = \begin{bmatrix} 20 \\ -11 \end{bmatrix}$, (c) $\mathbf{x} = \begin{bmatrix} 3 \\ -5 \end{bmatrix}$.

7. If $A^3 - 2A + 3I = 0$, then $(\frac{2}{3}I - \frac{1}{3}A^2)A = I$. So $\frac{2}{3}I - \frac{1}{3}A^2$ is the inverse of A.

8. (a) $E_{21}(3)$, $E_{21}(-3)$, (b) $E_2(-2)$, $E_2(-1/2)$

13. $(I + N + N^2 + \ldots + N^{k-1})(I - N) = (I + N + N^2 + \ldots + N^{k-1})I - (I + N + N^2 + \ldots + N^{k-1})N = (I + N + N^2 + \ldots + N^{k-1}) - (N + N^2 + N^3 + \ldots + N^k) = I - N^k$.

14. If N is nilpotent, $(I - N)^{-1} = I + N + N^2 + \ldots + N^{k-1}$.
(a) $\begin{bmatrix} 1 & -1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 1 & -3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$.

15. $f(x_3) \approx 0.7391$

Section 2.6, Page 104

2. (a) 3, (b) $1 + i$, (c) 1, (d) -70, (e) -6, (f) -6, (g) 1. All of the matrices are invertible.

3. $\det A = -5$, $\det B = 4$, and $\det AB = -20$ so $\det AB = \det A \det B$. Check that $\det A^T = -5$ and $\det B^T = 4$.

6 $\det V = (x_1 - x_0)(x_2 - x_0)(x_2 - x_1)$.

7. $\det A \det A^{-1} = \det AA^{-1} = \det I = 1$

8. Let $A = \begin{bmatrix} 1 & 1-i \\ 0 & 1+i \end{bmatrix}$. Then $\det A^H = 1 - i$ and $\det A = 1 + i$ so $\det A^H \neq \det A$.

In general, $\det A^H = \overline{\det A}$. By D7 and the definition of A^H, $\det A^H = \det \overline{A}$. Two facts about conjugates in section 1.2 are $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$ and $\overline{z_1 z_2} = \overline{z_1} \cdot \overline{z_2}$. Since the computation of $\det \overline{A}$ is a combination of taking the products and sums of conjugates of the coefficients of A, $\det A^H = \overline{\det A}$.

9. Let A be a square matrix with a non-zero determinant. By determinantal law D2, if you multiply one row of A by the scalar multiple zero, the determinant of the new matrix is $0 \cdot \det A = 0$ so any matrix with a row of zeros has zero determinant.

10. $-32\det(A)$, $\frac{1}{\det(A)}$

11. After preforming row operations on M, it can be reduced to $M' = \begin{bmatrix} \operatorname{rref}(A) & * \\ 0 & \operatorname{rref}(C) \end{bmatrix}$. After reduction, M', $\operatorname{rref}(A)$, and $\operatorname{rref}(C)$ are all upper triangular so $\det M' = \det(\operatorname{rref}(A))\det(\operatorname{rref}(C))$.
Let A' be the product of the row operations to reduce A and C' be the product of the row operations to reduce C. So the product of the row operations to reduce M is $A'C'$. So $\det M' = \det M \det(A'C') = \det M \det A' \det C'$ and $\det(\operatorname{rref}(A))\det(\operatorname{rref}(C)) = \det A \det A' \det C \det C'$. Since $\det M' = \det(\operatorname{rref}(A))\det(\operatorname{rref}(C))$, $\det M = \det A \det C$.

Section 2.7, Page 113

2. (a) $\begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -1 & 0 & 1 \end{bmatrix}$, (b) $\begin{bmatrix} -1 & 1 \\ 1 & -\frac{1}{2} \end{bmatrix}$

3. (b) $\begin{array}{l} x_1 = \frac{1}{4}b_1 + \frac{1}{4}b_2 \\ x_2 = \frac{1}{2}b_1 - \frac{1}{2}b_2 \end{array}$, (c) $\begin{array}{l} x_1 = \frac{-7}{6} \\ x_2 = \frac{5}{3} \\ x_3 = \frac{11}{2} \end{array}$

4. The system is $\begin{array}{l} y_0 = c_0 + c_1x_0 + c_2x_0^2 \\ y_1 = c_0 + c_1x_1 + c_2x_1^2 \\ y_2 = c_0 + c_1x_2 + c_2x_2^2. \end{array}$
The determinant of the coefficient matrix is $(x_1 - x_0)(x_2 - x_0)(x_2 - x_1)$ which equals 0 when $x_1 = x_0$, $x_1 = x_2$, or $x_2 = x_0$.

5. Since the matrix of minors of A has integer coefficients, $\operatorname{adj} A$ must be a matrix of integer coefficients. Since $A^{-1} = \frac{1}{\det A} \operatorname{adj} A = -1 \cdot \operatorname{adj} A$, A^{-1} is the product of an integer scalar and a matrix with integer coefficients so A^{-1} must have integer coefficients.

Section 2.8, Page 117

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

Section 2.9, Page 123

1. $L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 1 & 1 \end{bmatrix}$, $U = \begin{bmatrix} 2 & -1 & 1 \\ 0 & 4 & -3 \\ 0 & 0 & -1 \end{bmatrix}$, $\mathbf{x} = [1, -2, 2]^T$

2. Let $A = [a_{ij}]$, $L = \begin{bmatrix} 1 & 0 & 0 \\ & \ddots & 0 \\ * & & 1 \end{bmatrix}$, and $U = \begin{bmatrix} u_{11} & & * \\ & \ddots & \\ 0 & & u_{nn} \end{bmatrix}$ where $u_{ii} \neq 0$. So $a_{11} = u_{11}$, but $a_{11} = 0$ so $a_{11} \neq u_{11}$. Since there is no U, there is not an LU factorization of A.

Section 2.10, Page 123

1. $\begin{bmatrix} 4 & 1 & 5 \\ 1 & 6 & 0 \\ 5 & 0 & -4 \end{bmatrix}$

2. A is invertible if $k \neq 6$, and then $A^{-1} = \begin{bmatrix} \frac{-3}{k-6} & \frac{1}{k-6} \\ \frac{k}{k-6} & \frac{-2}{k-6} \end{bmatrix}$.

3. 21

4. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Then A and B are invertible matrices since $\det A \neq 0$ and $\det B \neq 0$, but $A + B$ is not invertible since $\det(A + B) = 0$.

5. Let $A = [a_{ij}]$ so the ij^{th} entry of $A + A^T$ is $a_{ij} + a_{ji}$ and the ji^{th} entry is $a_{ji} + a_{ij}$. So $A + A^T$ is symmetric.

Section 3.1, Page 134

1. V is a vector space.

2. V is not a vector space because it is not closed under vector addition and scalar multiplication.

3. V is not a vector space because it is not closed under scalar multiplication.

4. V is a vector space.

5. V is not a vector space because it is not closed under vector addition.

6. V is a vector space.

7.

$$\begin{aligned} c\mathbf{0} &= c(\mathbf{0}+\mathbf{0}) \\ c\mathbf{0} &= c\mathbf{0}+c\mathbf{0} \\ (c\mathbf{0}+(-c\mathbf{0})) &= c\mathbf{0}+(c\mathbf{0}+(-\mathbf{c0})) \\ \mathbf{0} &= c\mathbf{0}+\mathbf{0} \\ \mathbf{0} &= c\mathbf{0} \end{aligned}$$

8.
$$\begin{aligned} (-c)\mathbf{v} &= (-1c)\mathbf{v} \\ &= -1(c\mathbf{v}) \\ &= -(c\mathbf{v}) \end{aligned}$$

Similarly, you can show $(-c)\mathbf{v} = c(-\mathbf{v}) = -(c\mathbf{v})$.

9. Suppose $c\mathbf{v} = \mathbf{0}$. If $c = 0$, we're done by Law 1. Else if $c \neq 0$, then there is $\frac{1}{c}$ so that

$$\begin{aligned} (\frac{1}{c})c\mathbf{v} &= (\frac{1}{c})\mathbf{0} \\ (\frac{1}{c}c)\mathbf{v} &= \mathbf{0} \\ 1\mathbf{v} &= \mathbf{0} \\ \mathbf{v} &= \mathbf{0} \end{aligned}$$

So $c = 0$ or $\mathbf{v} = \mathbf{0}$.

13. (a) Is linear, but range is not V, (b) Is not linear, (c) Is linear, range is V.

14. (a) T is not a linear transformation because $T((1,0)+(0,1)) = T(1,1) = 1(0,1) = (0,1)$ but $T(1,0)+T(0,1) = 1(0,0)+0(0,1) = (0,0)$.

Section 3.2, Page 141

1. W is not a subspace of V because W is not closed under addition and scalar multiplication.

2. W is a subspace of V.

3. W is a subspace of V.

4. W is a subspace of V.

5. W is a subspace of V.

6. W is not a subspace of V because W is not closed under scalar multiplication.

7.Not a subspace, since W doesn't contain the zero element.

11. (a) Spans $\mathcal{P}_2$, (b) does not span $\mathcal{P}_2$

12. For example, $\mathbf{v} = (1,1,0)$ and $\mathbf{w} = (1,-1,-2)$.

13. The spans are equal.

14. Let A and B be $n \times n$ diagonal matrices. Then cA is a diagonal matrix and $A+B$ is a diagonal matrix so the set of diagonal matrices is closed under matrix addition and scalar multiplication. Therefore the set of diagonal matrices is a subspace of $\mathbb{R}^{n,n}$.

15. (a) If $x,y \in U$ and $x,y \in V$, $x,y \in U \cap V$. Then $cx \in U$ and $cx \in V$ so $cx \in U \cap V$, and $x+y \in U$ and $x+y \in V$ so $x+y \in U \cap V$. Therefore $U \cap V$ is closed under addition and scalar multiplication so it is a subspace of W.
(b) If $u_1,u_2 \in U$ and $v_1,v_2 \in V$, $u_1+v_1, u_2+v_2 \in U+V$. Then $cu_1 \in U$ and $cv_1 \in V$ so $c(u_1+v_1) = cu_1+cv_1 \in U+V$, and $u_1+u_2 \in U$ and $v_1+v_2 \in V$ so $(u_1+v_1)+(u_2+v_2) = (u_1+u_2)+(v_1+v_2) \in U+V$. Therefore $U+V$ is closed under addition and scalar multiplication so it is a subspace of W.

17. (a) If $A = [a_{ij}]$, $\mathrm{vec}(A) = (a_{11},a_{21},a_{12},a_{22})$ so for A there exists only one $\mathrm{vec}(A)$. If $\mathrm{vec}(A) = (a_{11},a_{21},a_{12},a_{22})$, $A = [a_{ij}]$ so for $\mathrm{vec}(A)$ there exists only one A. Therefore the vec

operation establishes a one-to-one correspondence between matrices in V and vectors in $\mathbb{R}^4$.
(b) Let $A = [a_{ij}]$ and $B = [b_{ij}]$. So $c\,\mathrm{vec}(A) + d\,\mathrm{vec}(B) = c(a_{11}, a_{21}, a_{12}, a_{22}) + d(b_{11}, b_{21}, b_{12}, b_{22}) = (ca_{11} + db_{11}, ca_{21} + db_{21}, ca_{12} + db_{12}, ca_{22} + db_{22}) = \mathrm{vec}(cA + dB)$.

18. If $k \geq 1$, $V_k = \mathrm{span}\{A^0, A^1\}$.

Section 3.3, Page 150

1. (a) linearly independent, (b) each vector is redundant, (c) each vector is redundant

3. (a) linearly independent, (b) linearly independent, (c) each vector is redundant

4. (a) $(-1, -2, 3)$, (b) $(\frac{1}{2}, 1, \frac{3}{2})$

5. (a) $\mathbf{v} \in \mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2\}$, (b) $\mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2, (1, 0, -1)\}$

6. (a) $\mathbf{v} \notin \mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2\}$, (b) $\mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}\}$

7. $\mathrm{span}\,\{v_1, v_2\}$

13. Start with a nontrivial linear combination of the functions that sums to 0 and differentiate it.

15. Assume $\mathbf{v}_i = \mathbf{v}_j$. Then there exists $c_i = -c_j \neq 0$ such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_i\mathbf{v}_i + \ldots + c_j\mathbf{v}_j + \ldots + c_n\mathbf{v}_n = \mathbf{0}$. Therefore $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly dependent.

Section 3.4, Page 159

1 (a) $\begin{bmatrix} 1 & -1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, (b) $\mathcal{N}(A) = \mathrm{span}\{\}$, (c) $\mathcal{C}(A) = \mathrm{span}\{(1,1,3), (0,1,2)\}$, (d) $\mathcal{R}(A) = \mathrm{span}\{(1,2,0,0,1), (1,2,1,1,1)\}$, (e) vector of coefficients belong to $\mathrm{span}\{(-2,1,0,0,0), (0,0,-1,1,0), (-1,0,0,0,1)\}$

2 (a) $\begin{bmatrix} 1 & 2 & 0 & 0 & 1 & 7 \\ 0 & 0 & 1 & 1 & 0 & \alpha - 7 \\ 0 & 0 & 0 & 0 & 0 & \beta - 2\alpha - 7 \end{bmatrix}$, (b) $\mathcal{N}(A) = \mathrm{span}\{(-2,1,0,0,0), (0,0,-1,1,0), (-1,0,0,0,1)\}$, (c) $\mathcal{C}(A) = \mathrm{span}\{(1,1,3), (0,1,2)\}$, (d) $\mathcal{R}(A) = \mathrm{span}\{(1,2,0,0,1), (1,2,1,1,1)\}$, (e) vector of coefficients belong to $\mathrm{span}\{(-2,1,0,0,0), (0,0,-1,1,0), (-1,0,0,0,1)\}$, (f) $\{(\alpha,\beta) | \beta - 2\alpha - 7 = 0\}$

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

6. T is not one-to-one since $\ker T \neq \{\mathbf{0}\}$.

10. Since A is nilpotent, there exists m such that $A^m = \mathbf{0}$ so $\det(A^m) = (\det A)^m = 0$ and $\det A = 0$. Therefore A is not invertible so $\mathcal{N}(A) \neq \{\mathbf{0}\}$. Also since A is nilpotent, by Exercise 14 in Section 2.4, $(I - A)^{-1} = I + A + A^2 + \ldots + A^{m-1}$. Since $I - A$ is invertible, $\mathcal{N}(I - A) = \{\mathbf{0}\}$.

Section 3.5, Page 164

1 (a) none

5. (a) true, (b) false, (c) true, (d) true, (e) true, (f) true, (g) false, (h) false, (i) false, (j) false, (k) true

7. Since $\mathcal{P}$ is a subspace of $C[0,1]$ and $\dim \mathcal{P}$ is infinite, $\dim(C[0,1])$ is infinite.

8. $\dim(\{E_{i,j}\}) = mn = \dim(\mathbb{R}^{m,n})$. If $c_{1,1}E_{1,1} + \ldots + c_{i,j}E_{i,j} = \mathbf{0}$, $c_{a,b} = 0$ for each a, b because $E_{a,b}$ is the only matrix with a nonzero entry in the (a,b)th position. Therefore $\{E_{i,j} \mid i = 1, \ldots, m,\ j = 1, \ldots, n\}$ is a basis of the vector space $\mathbb{R}^{m,n}$.

13. The dimension of the space is $n(n+1)/2$.

Section 3.6, Page 173

1. $\mathcal{R}(A) = \mathrm{span}\{(1,0), (0,1)\}$, $\mathcal{C}(A) = \mathrm{span}\{(0,-1,1), (2,1,1)\}$, $\mathcal{N}(A) =$

$\operatorname{span}\{\mathbf{0}\}$,
$\mathcal{R}(B) = \operatorname{span}\{(1,0,-2),(0,1,0)\}$,
$\mathcal{C}(B) = \operatorname{span}\{(2,-1,1),(2,0,1)\}$,
$\mathcal{N}(B) = \operatorname{span}\{(2,0,1)\}$,
$\mathcal{R}(C) = \operatorname{span}\{(1,0,-2,-2),(0,1,\frac{1}{2},2)\}$,
$\mathcal{C}(C) = \operatorname{span}\{(1,-1),(2,0)\}$, $\mathcal{N}(C) = \operatorname{span}\{(2,\frac{-1}{2},1,0),(2,-2,0,1)\}$

2. (a) (c_1,c_2,c_3,c_4) such that $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3+c_4\mathbf{v}_4=\mathbf{0}$ where $c_1=-2c_3-2c_4$, $c_2=-c_3$, and c_3,c_4 are free. Dimension of $\operatorname{span}\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4\}$ is 2.
(b) (c_1,c_2,c_3) such that

$c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=\mathbf{0}$ where $c_1=2c_3$, $c_2=-c_3$, and c_3 is free.

4. (a) $\mathcal{R}(A) = \operatorname{span}\{(1,0,-1,0,0,-2,-3), (0,1,1,0,0,2,3), (0,0,0,1,0,4,5), (0,0,0,0,1,6,7)\}$, $\operatorname{rank} A = 4$. $\mathcal{C}(A) = \operatorname{span}\{(3,1,3,0,0), (1,1,2,2,3), (0,-1,1,-1,3), (1,1,1,1,-3)\}$. $\mathcal{N}(A) = \operatorname{span}\{(1,-1,1,0,0,0,0), (2,-2,0,-4,-6,1,0), (3,-3,0,-5,-7,0,1)\}$, $nullity(A)=3$.

10. Since $A\mathbf{x}=\mathbf{b}$ is a consistent, $\mathbf{b}\in\mathcal{C}(A)$. If $\{\mathbf{c}_i\}$, the set of columns of A, has redundant vectors in it, $a_1\mathbf{c}_1+a_2\mathbf{c}_2+\ldots+a_n\mathbf{c}_n=\mathbf{0}$ for some $a_i\neq 0$. If $d_1\mathbf{c}_1+d_2\mathbf{c}_2+\ldots+d_n\mathbf{c}_n=\mathbf{b}$ is a solution for $A\mathbf{x}=\mathbf{b}$, then $(a_1+d_1)\mathbf{c}_1+(a_2+d_2)\mathbf{c}_2+\ldots+(a_n+d_n)\mathbf{c}_n=\mathbf{b}$ is also a solution. Therefore the system has more than one solution.

11. $\dim(M_{n\times n}(\mathbb{R})) = n^2$ so $\{I,A,A^2,\ldots,A^{n^2}\}$ must be linearly dependent since $\dim\{I,A,A^2,\ldots,A^{n^2}\} = n^2+1 > \dim(M_{n\times n}(\mathbb{R}))$. So there exists $c_0I+c_1A+c_2A^2+\cdots+c_{n^2}A^{n^2}=\mathbf{0}\in M_{n\times n}(\mathbb{R})$ with some $c_i\neq 0$. Pick $p(x)=c_0+c_1x+\cdots+c_{n^2}A^{n^2}$ so $p(x)\neq\mathbf{0}$ and $p(A)=0$.

Section 3.7, Page 178

1. $\begin{bmatrix} 1 & 2 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$, $\operatorname{dom}(T) = \operatorname{span}\{(1,0,0),(0,1,0),(0,0,1)\}$, $\operatorname{range}(T) = \operatorname{span}\{(1,1,0),(2,-1,1),(0,0,1)\}$, $\ker(T)=\{\mathbf{0}\}$

Section 3.9, Page 182

1. (a) W is not a subspace of V because W is not closed under matrix addition. (b) W is not a subspace of V because W is not closed under matrix addition.

9. (a) true, (b) false, (c) true, (d) true, (e) true, (f) false, (g) false, (h) false, (i) false (consider $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$).

Section 4.1, Page 191

1. (c) $\sqrt{10}$, $2+i$

6. $\|c\mathbf{v}\| = |c|\,\|\mathbf{v}\|$ by Basic Norm Law #2. Since $c\in\mathbb{R}$ and $c>0$, $\|c\mathbf{v}\| = c\,\|\mathbf{v}\|$. So a unit vector in direction of $c\mathbf{v}$ is $\mathbf{u}_{c\mathbf{v}} = c\mathbf{v}/c\,\|\mathbf{v}\| = \mathbf{v}/\|\mathbf{v}\| = \mathbf{u}_{\mathbf{v}}$. If $c<0$, then $\mathbf{u}_{c\mathbf{v}} = -\mathbf{u}_{\mathbf{v}}$.

7. Let $\mathbf{u}=(u_1,\ldots,u_n)\in\mathbb{R}^n$, $\mathbf{v}=(v_1,\ldots,v_n)\in\mathbb{R}^n$, and $c\in\mathbb{R}$. Then $(c\mathbf{u})\cdot\mathbf{v} = (cu_1)v_1+\ldots+(cu_n)v_n$ and $\mathbf{v}\cdot(c\mathbf{u}) = v_1(cu_1)+\ldots+v_n(cu_n)$ so $(c\mathbf{u})\cdot\mathbf{v} = \mathbf{v}\cdot(c\mathbf{u})$. Similarly, you can show $(c\mathbf{u})\cdot\mathbf{v} = \mathbf{v}\cdot(c\mathbf{u}) = c(\mathbf{v}\cdot\mathbf{u}) = c(\mathbf{u}\cdot\mathbf{v})$.

Section 4.2, Page 200

2. (a) $90\,°$, (b) $\mathbf{u}_{\mathbf{v}_1} = [\frac{\sqrt{2}}{2},0,\frac{\sqrt{2}}{2}]^T$, $\mathbf{u}_{\mathbf{v}_2} = [\frac{\sqrt{3}}{3},\frac{\sqrt{3}}{3},-\frac{\sqrt{3}}{3}]^T$, (c) $\mathbf{0}$, (d) $|\mathbf{v}_1\cdot\mathbf{v}_2| = 0$ and $\|\mathbf{v}_1\|\,\|\mathbf{v}_2\| = \sqrt{6}$ so $|\mathbf{v}_1\cdot\mathbf{v}_2| \leq \|\mathbf{v}_1\|\,\|\mathbf{v}_2\|$

4. The normal equations are $A^TA\mathbf{x} = A^T\mathbf{b}$.
(1) $\mathbf{x} = \left[\frac{3}{7}, \frac{-2}{3}\right]^T$, $\mathbf{b} - A\mathbf{x} = \left[\frac{-4}{21}, \frac{-16}{21}, \frac{8}{21}\right]^T$, $||\mathbf{b} - A\mathbf{x}|| = \frac{4\sqrt{21}}{21}$

(2) $\mathbf{x} = \left[\frac{9}{7} - x_3, \frac{4}{7} - x_3, x_3\right]^T$ where x_3 is free, $\mathbf{b} - A\mathbf{x} = \left[\frac{2}{7}, \frac{-6}{7}, \frac{4}{7}\right]$, $||\mathbf{b} - A\mathbf{x}|| = \frac{2\sqrt{14}}{7}$

Section 4.3, Page 208

1. (a),(b), (c), (d) are linearly independent; (a), (c) and (d) are orthogonal; (c) is an orthonormal set.

2. $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, $\mathbf{v}_1 \cdot \mathbf{v}_3 = 0$, and $\mathbf{v}_2 \cdot \mathbf{v}_3 = 0$ so $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis of $\mathbb{R}^3$.

The coordinates of $\mathbf{v}$, (c_1, c_2, c_3), such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{v}$ are $(\frac{3}{2}, \frac{-1}{3}, \frac{-5}{3})$.

3. $\{\theta = \pi k | k \in \mathbb{Z}\}$ makes A symmetric.

Section 4.5, Page 212

1. $[\frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6}, \frac{-\sqrt{6}}{3}]^T$

2. $||\mathbf{u}|| = \sqrt{7}$, $\mathbf{u} \cdot \mathbf{v} = 0$

Section 5.1, Page 222

1. (a) $-2, -1$, (b) $2, 0, 5$, (c) $2, 2, 3$, (e) $-1, -1, -1$.

2. (a) The basis of $\mathcal{E}_{-2}$ is $\{(3, 1)\}$, and the basis of $\mathcal{E}_{-1}$ is $\{(4, 1)\}$. Both algebraic and geometric multiplicity of each is 1. (b) The basis of $\mathcal{E}_2$ is $\{(1, 0, 0)\}$, the basis of $\mathcal{E}_0$ is $\{(0, 1, -3)\}$, and the basis of $\mathcal{E}_5$ is $\{(0, 1, 2)\}$. (c) The basis of $\mathcal{E}_2$ is $\{(0, -1, 1), (1, 0, 0)\}$, and the basis of $\mathcal{E}_3$ is $\{(1, 1, 0)\}$. (e) A basis of $\mathcal{E}_{-1}$ is $\{(0, 0, 1)\}$. The algebraic multiplicity of the eigenvalue -1 is 3, while the geometric multiplicity is 1.

4. (a) $\operatorname{tr} A = -3 = -2 + -1$, (b) $\operatorname{tr} A = 7 = 2 + 0 + 5$, (c) $\operatorname{tr} A = 7 = 2 + 2 + 3$

5. (a) $p(\lambda) = (a - \lambda)(d - \lambda) - bc$, $\lambda_1 = \frac{1}{2}a + \frac{1}{2}d + \frac{1}{2}\sqrt{(a^2 - 2ad + d^2 + 4bc)}$, $\lambda_2 = \frac{1}{2}a + \frac{1}{2}d - \frac{1}{2}\sqrt{(a^2 - 2ad + d^2 + 4bc)}$, (b) $\operatorname{tr} A = a + d = \lambda_1 + \lambda_2$, (c) $\det A = ad - bc = \lambda_1\lambda_2$

6. For A, the basis of $\mathcal{E}_1$ is $\{(1, 0)\}$, and the basis of $\mathcal{E}_2$ is $\{(1, 1)\}$. For A^T, the basis of $\mathcal{E}_1$ is $\{(-1, 1)\}$, and the basis of $\mathcal{E}_2$ is $\{(0, 1)\}$.

7. The eigenvalues of A are $\frac{3}{2} + \frac{\sqrt{3}}{2}i$ and $\frac{3}{2} - \frac{\sqrt{3}}{2}i$.

8. Since $(\lambda I - A)^T = \lambda I - A^T$ and $\det(\lambda I - A)^T = \det(\lambda I - A)$, $\det(\lambda I - A) = \det(\lambda I - A^T)$ so A and A^T have the same eigenvalues.

9. Since $\mathbf{x}$ is an eigenvector of A with eigenvalue λ, $A\mathbf{x} = \lambda\mathbf{x}$. So $A(c\mathbf{x}) = c(A\mathbf{x}) = c(\lambda\mathbf{x}) = \lambda(c\mathbf{x})$. Therefore $c\mathbf{x}$ is an eigenvector of A with eigenvalue λ.

10. Let λ be an eigenvalue of A with eigenvector $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$. So $(I - A)\mathbf{v} = I\mathbf{v} - A\mathbf{v} = \mathbf{v} - \lambda\mathbf{v} = (1 - \lambda)\mathbf{v}$. Thus if λ is an eigenvalue of A, $1 - \lambda$ is an eigenvalue of $I - A$. Since $|\lambda| < 1$, $1 - \lambda > 0$ so every eigenvalue of $I - A$ is nonzero. Therefore $I - A$ is invertible.

11. Let λ be an eigenvalue of A with eigenvector $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$. If A is invertible, $\lambda \neq 0$ so $A^{-1}A\mathbf{v} = A^{-1}\lambda\mathbf{v}$ and thus $A^{-1}\mathbf{v} = \frac{1}{\lambda}\mathbf{v}$. Therefore $1/\lambda$ is an eigenvalue of A^{-1}.

Section 5.2, Page 230

7. (a) $A\mathbf{x}_n = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} f_{n+1} \\ f_n \end{bmatrix} = \begin{bmatrix} f_{n+1} + f_n \\ f_{n+1} \end{bmatrix} = \begin{bmatrix} f_{n+2} \\ f_{n+1} \end{bmatrix} = \mathbf{x}_{n+1}$

(b) $f_n = (\frac{1+\sqrt{5}}{2})^n(\frac{5+\sqrt{5}}{10}) + (\frac{1-\sqrt{5}}{2})^n(\frac{5-\sqrt{5}}{10})$

Section 5.3, Page 238

1. (a), (b), (c), and (e) are diagonalizable because they are non-defective. (d) and (f) are not diagonalizable because they are defective.

2. (a) $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, (b) $P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$, $D = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{-1}{2} \end{bmatrix}$, (c) $P = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$, $D = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{-1}{2} \end{bmatrix}$

3. (b) has $\lim A^k = \mathbf{0}$.

6. (a) there is no dominant eigenvalue, (b) -4, (c) there is no dominant eigenvalue

Section 5.4, Page 243

1. (a) $2+\sqrt{2}, 2-\sqrt{2}$, (b) $-3, 2$, (c) $0, 2, 3$,(f) $1, 1 \pm 2\sqrt{2}$

2. (a) The basis of $\mathcal{E}_{2+\sqrt{2}}$ is $\{(-i - i\sqrt{2}, 1)\}$, and the basis of $\mathcal{E}_{2-\sqrt{2}}$ is $\{(1, -i + i\sqrt{2})\}$. (b) The basis of $\mathcal{E}_{-3}$ is $\{(-2, 1)\}$, and the basis of $\mathcal{E}_2$ is $\{(1, 2)\}$. Also $P = \frac{\sqrt{5}}{5}\begin{bmatrix} -2 & 1 \\ 1 & 2 \end{bmatrix}$, $D = \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix}$ (c) The basis of $\mathcal{E}_0$ is $\{(-1-i, 1, 0)\}$, the basis of $\mathcal{E}_2$ is $\{(0, 0, 1)\}$, and the basis of $\mathcal{E}_3$ is $\{(1, 1 - i, 0)\}$. (d) The basis of $\mathcal{E}_0$ is $\{(1, \frac{-1}{2} + \frac{i}{2})\}$, and the basis of $\mathcal{E}_3$ is $\{(\frac{1}{2}+\frac{i}{2}, 1)\}$. Also $P = \frac{\sqrt{6}}{6}\begin{bmatrix} 2 & 1+i \\ -1+i & 2 \end{bmatrix}$, $D = \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}$.

4. (a) $A = A^H$, (b) the basis of $\mathcal{E}_2$ is $\{(-i, 1)\}$, the basis of $\mathcal{E}_4$ is $\{(i, 1)\}$, (c) $P = \frac{\sqrt{2}}{2}\begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$, (d) $A^k = PD^kP^T$

Section 5.7, Page 258

3. Eigenvalues are $5, -4, 1 + 2i\sqrt{5}, 1 - 2i\sqrt{5}$.

Section 5.8, Page 259

2. $P = \begin{bmatrix} 0 & 1 & 3 \\ 1 & 0 & 4 \\ -1 & 0 & 0 \end{bmatrix}$, $D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

4. (a) false, (b) true, (c) false, (d) false, (e) true

5. Let λ be an eigenvalue of A and $\mathbf{v}$ be an eigenvector of A so that $A\mathbf{v} = \lambda\mathbf{v}$. So $(A+I)\mathbf{v} = A\mathbf{v} + I\mathbf{v} = \lambda\mathbf{v} + \mathbf{v} = (\lambda+1)\mathbf{v}$. Therefore, $\lambda + 1$ is an eigenvalue of $A + I$.

7. The eigenvalues of A are $\mathbf{a} \cdot \mathbf{b}$ and 0 (with multiplicity $n - 1$). The basis of $\mathcal{E}_{\mathbf{a}\cdot\mathbf{b}}$ is $\{\mathbf{a}\}$, and the basis of $\mathcal{E}_0$ is $\{(-b_2, b_1, 0, \ldots, 0), (-b_3, 0, b_1, 0, \ldots, 0), \ldots, (-b_n, 0, \ldots, 0, b_1)\}$.

9. Let A be Hermitian symmetric so $A = A^H$. So $A(A^H) = A^H(A^H) = A^H(A)$. Therefore A is *normal*.

Section 6.1, Page 265

1. $\|\mathbf{x}\|_1 = 2+\sqrt{2}$, $\|\mathbf{x}\|_2 = 2$, $\|\mathbf{x}\|_\infty = \sqrt{2}$, $\|\mathbf{y}\|_1 = 8$, $\|\mathbf{y}\|_2 = \sqrt{22}$, $\|\mathbf{y}\|_\infty = 4$

2. $\mathbf{u}_1 = \frac{1}{5}(1,-3,-1)$, $\mathbf{u}_2 = \frac{\sqrt{11}}{11}(1,-3,-1)$, $\mathbf{u}_\infty = \frac{1}{3}(1,-3,-1)$

3. $\|\mathbf{u}+\mathbf{v}\|_1 = 7 \le 6+7 = \|\mathbf{u}\|_1 + \|\mathbf{v}\|_1$

5. $\sqrt{2}$

6. $\max\{||A\mathbf{x}|| \mid ||\mathbf{x}|| = 1\} = \max\{|a|+|b|, |c|+|d|\}$

7. Let $\mathbf{u} = (u_1, \dots u_n)$ and $\mathbf{v} = (v_1, \dots v_n)$ so $\|\mathbf{u}\|_1 = |u_1| + \dots + |u_n| \ge 0$. Also $\|c\mathbf{u}\|_1 = |cu_1| + \dots + |cu_n| = |c|\,|u_1| + \dots + |c|\,|u_n| = |c|\,\|\mathbf{u}\|_1$, and $\|\mathbf{u}+\mathbf{v}\|_1 = |u_1+v_1| + \dots + |u_n+v_n| \le |u_1| + \dots + |u_n| + |v_1| + \dots + |v_n| = \|\mathbf{u}\|_1 + \|\mathbf{v}\|_1$.

8. $\mathbf{v} = \lim_{n\to\infty} \mathbf{v}_n = [-1,1]^T$. $\mathbf{v} - \mathbf{v}_n = \left[\frac{-1}{n}, -e^{-n}\right]^T$ so $\|\mathbf{v}-\mathbf{v}_n\|_1 = (\frac{1}{n} + e^{-n})\xrightarrow[n\to\infty]{} 0$ and $\|\mathbf{v}-\mathbf{v}_n\|_2 = \sqrt{(\frac{1}{n})^2 + (e^{-n})^2} \xrightarrow[n\to\infty]{} 0$. Therefore $\lim_{n\to\infty}\mathbf{v}_n$ is the same with respect to both that 1- and 2-norm.

9. $[0,0]^T$

Section 6.2, Page 274

1. $|< \mathbf{u}, \mathbf{v} >| = 4 \le \sqrt{70} = \|\mathbf{u}\|\,\|\mathbf{v}\|$

2. (a) $\theta \approx 118.56\,^\circ$, (b) $\theta \approx 23.58\,^\circ$

3. All are linearly independent; (a), (c), (d) and (e) are orthogonal; (c) is an orthonormal set.

4. $< [x_1,x_2]^T, [x_1,x_2]^T > = 3x_1^2 - 2x_2^2$. Since $3x_1^2 - 2x_2^2$ is not necessarily greater than or equal to 0, the given law fails the first condition of the definition of an inner product.

7.
$$\begin{aligned} < \mathbf{u}, \mathbf{v} > &= (A\mathbf{u})^H A\mathbf{v} \\ &= [2u_1, 3u_2]\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ &= 2u_1v_1 + 3u_2v_2 \end{aligned}$$

8. $\frac{\sqrt{2}}{10}$

9. If $\mathbf{w} \in V$, then Theorem 4.3.3 supplies a formula for $\mathbf{w}$, which we can check. Doing so shows $\mathbf{w} \notin \operatorname{span}\{(1,1,0),(-1,1,1)\}$.

Section 6.3, Page 284

1. For standard inner product $(\frac{5}{2}, \frac{1}{2}, 1)$; the other gives $(\frac{19}{10}, \frac{9}{10}, \frac{1}{2})$

2. (a) $\{(1,0,0), \frac{\sqrt{2}}{2}(0,-1,-1), \frac{\sqrt{2}}{2}(0,-1,1)\}$, (b) $(2, \frac{5}{2}, \frac{5}{2})$, (c) $(2, \frac{5}{2}, \frac{5}{2}) + (0, \frac{-3}{2}, \frac{3}{2})$, (d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, (e) $P\mathbf{w} = \left[2, \frac{5}{2}, \frac{5}{2}\right]^T$

3. (a) $V = \operatorname{span}\{\frac{1}{2}(-1,-1,1,1), \frac{1}{2}(1,1,1,1), \frac{\sqrt{2}}{2}(0,0,-1,1)\}$, (b) $(1/2, 1/2, 0, 0)$, (c) $(1/2,1/2,0,0) + (\frac{1}{7}2, -1/2, 0, 0)$, (d) $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, (e) $P\mathbf{w} = (\frac{1}{2}, \frac{1}{2}, 0, 0)$

4. (a) $\{\frac{\sqrt{3}}{3}(1,1,-1), \frac{\sqrt{78}}{78}(-2,7,5)\}$

Section 6.4, Page 293

1. (a) $V^\perp = \operatorname{span}\{(1,1,0),(-2,0,1)\}$

3. $V^\perp = \operatorname{span}\{(\frac{3}{2},1,0),(-1,0,1)\}$

10. $U \cap V = \operatorname{span}\{(3,3,1)\}$.

Section 6.7, Page 306

1. (a) $\{(1,1,1,1),(1,-1,1,-1),(0,-1,0,1)\}$, (b) $\{\frac{1}{2}(1,1,1,1), \frac{1}{2}(1,-1,1,-1), \frac{\sqrt{2}}{2}(0,-1,0,1)\}$, (c) $\begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

2. $\frac{\sqrt{2}}{10}$

3. The formula does not define an inner product on $\mathbb{R}^2$.

Bibliography

[1] Richarc Bellman. *Introduction to Matrix Analysis, 2nd ed.* SIAM, Philadelphia, PA, 1997.

[2] G. Caughley. Parameters for seasonally breeding populations. *Ecology*, 48:834–839, 1967.

[3] Biswa Nath Datta. *Numerical Linear Algebra and Applications*. Brooks/Cole, New York, 1995.

[4] James W. Demmel. *Applied Numerical Linear Algebra*. SIAM, Philadelphia, PA, 1997.

[5] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, Maryland, 1983.

[6] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985.

[7] P. Lancaster and M. Tismenetsky. *The Theory of Matrices*. Academic Press, Orlando, Florida, 1985.

[8] Lloyd Trefethen and David Bau. *Numerical Linear Algebra*. SIAM, Philadelphia, PA, 1997.

Index