Linear Algebra in Twenty Five Lectures

Tom Denton and Andrew Waldron

August 4, 2014

Edited by Rohit Thomas

Contents

Preface

These linear algebra lecture notes are designed to be presented as twenty five, fifty minute lectures suitable for sophomores likely to use the material for applications but still requiring a solid foundation in this fundamental branch of mathematics. The main idea of the course is to emphasize the concepts of vector spaces and linear transformations as mathematical structures that can be used to model the world around us. Once "persuaded" of this truth, students learn explicit skills such as Gaussian elimination and diagonalization in order that vectors and linear transformations become calculational tools, rather than abstract mathematics.

In practical terms, the course aims to produce students who can perform computations with large linear systems while at the same time understand the concepts behind these techniques. Often-times when a problem can be reduced to one of linear algebra it is "solved". These notes do not devote much space to applications (there are already a plethora of textbooks with titles involving some permutation of the words "linear", "algebra" and "applications"). Instead, they attempt to explain the fundamental concepts carefully enough that students will realize for their own selves when the particular application they encounter in future studies is ripe for a solution via linear algebra.

The notes are designed to be used in conjunction with a set of online homework exercises which help the students read the lecture notes and learn basic linear algebra skills. Interspersed among the lecture notes are links to simple online problems that test whether students are actively reading the notes. In addition there are two sets of sample midterm problems with solutions as well as a sample final exam. There are also a set of ten online assignments which are collected weekly. The first assignment is designed to ensure familiarity with some basic mathematic notions (sets, functions, logical quantifiers and basic methods of proof). The remaining nine assignments are devoted to the usual matrix and vector gymnastics expected from any sophomore linear algebra class. These exercises are all available at

https://ftcourses.webwork.maa.org/webwork2/ft-marlboro-nsc164/

Webwork is an open source, online homework system which originated at the University of Rochester. It can efficiently check whether a student has answered an explicit, typically computation-based, problem correctly. The problem sets chosen to accompany these notes could contribute roughly 20% of a student's grade, and ensure that basic computational skills are mastered. Most students rapidly realize that it is best to print out the Webwork assignments and solve them on paper before entering the answers online. Those who do not tend to fare poorly on midterm examinations. We have found that there tend to be relatively few questions from students in office hours about the Webwork assignments. Instead, by assigning 20% of the grade to written assignments drawn from problems chosen randomly from the review exercises at the end of each lecture, the student's focus was primarily on understanding ideas. They range from simple tests of understanding of the material in the lectures to more difficult problems, all of them require thinking, rather than blind application of mathematical "recipes". Office hour questions reflected this and offered an excellent chance to give students tips how to present written answers in a way that would convince the person grading their work that they deserved full credit!

Each lecture concludes with references to the comprehensive online textbooks of Jim Hefferon and Rob Beezer:

http://joshua.smcvt.edu/linearalgebra/

http://linear.ups.edu/index.html

and the notes are also hyperlinked to Wikipedia where students can rapidly access further details and background material for many of the concepts. Videos of linear algebra lectures are available online from at least two sources:

- The Khan Academy,
  http://www.khanacademy.org/?video#Linear Algebra
- MIT OpenCourseWare, Professor Gilbert Strang,
  http://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/video-lectures/

There are also an array of useful commercially available texts. A non-exhaustive list includes

- "Introductory Linear Algebra, An Applied First Course", B. Kolman and D. Hill, Pearson 2001.
- "Linear Algebra and Its Applications", David C. Lay, Addison–Weseley 2011.

- "Introduction to Linear Algebra", Gilbert Strang, Wellesley Cambridge Press 2009.
- "Linear Algebra Done Right", S. Axler, Springer 1997.
- "Algebra and Geometry", D. Holten and J. Lloyd, CBRC, 1978.
- "Schaum's Outline of Linear Algebra", S. Lipschutz and M. Lipson, McGraw-Hill 2008.

A good strategy is to find your favorite among these in the University Library.

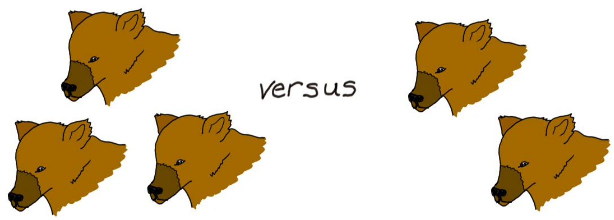
There are still many errors in the notes, as well as awkwardly explained concepts. An army of 400 students, Fu Liu, Stephen Pon and Gerry Puckett have already found many of them. Rohit Thomas has spent a great deal of time editing these notes and has improved them immeasurably. We also thank Captain Conundrum for providing us his solutions to the sample midterm and final questions. The review exercises would provide a better survey of what linear algebra really is if there were more "applied" questions. We welcome your contributions!

Andrew and Tom.

1 What is Linear Algebra?

Three bears go into a cave, two come out. Would you go in?

Brian Butterworth



Numbers are highly useful tools for surviving in the modern world, so much so that we often introduce abstract *pronumerals* to represent them:

n bears go into a cave, $n-1$ come out. Would you go in?

A single number alone is not sufficient to model more complicated real world situations. For example, suppose I asked everybody in this room to rate the likeability of everybody else on a scale from 1 to 10. In a room full of n people (or bears *sic*) there would be n^2 ratings to keep track of (how much Jill likes Jill, how much does Jill like Andrew, how much does Andrew like Jill, how much does Andrew like Andrew, *etcetera*). We could arrange these in a square array

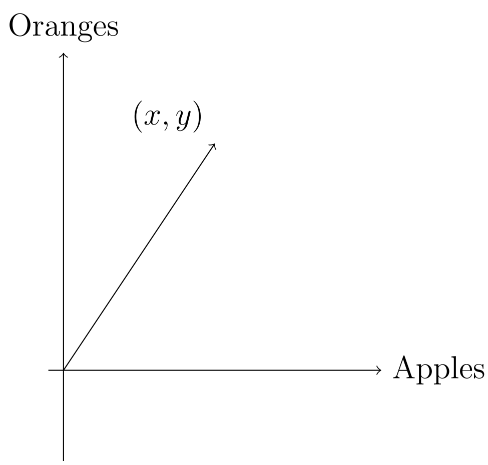
$$\begin{pmatrix} 9 & 4 & \cdots \\ 10 & 6 & \\ \vdots & & \ddots \end{pmatrix}$$

Would it make sense to replace such an array by an abstract symbol M? In the case of numbers, the pronumeral n was more than a placeholder for a particular piece of information; there exists a myriad of mathematical operations (addition, subtraction, multiplication,...) that can be performed with the symbol n that could provide useful information about the real world system at hand. The array M is often called a *matrix* and is an example of a more general abstract structure called a *linear transformation* on which many mathematical operations can also be defined. (To understand why having an abstract theory of linear transformations might be incredibly useful and even lucrative, try replacing "likeability ratings" with the number of times internet websites link to one another!) In this course, we'll learn about three main topics: Linear Systems, Vector Spaces, and Linear Transformations. Along the way we'll learn about matrices and how to manipulate them.

For now, we'll illustrate some of the basic ideas of the course in the two dimensional case. We'll see everything carefully defined later and start with some simple examples to get an idea of the things we'll be working with.

Example Suppose I have a bunch of apples and oranges. Let x be the number of apples I have, and y be the number of oranges I have. As everyone knows, apples and oranges don't mix, so if I want to keep track of the number of apples and oranges I have, I should put them in a list. We'll call this list a *vector,* and write it like this: (x, y). The order here matters! I should remember to always write the number of apples first and then the number of oranges - otherwise if I see the vector $(1, 2)$, I won't know whether I have two apples or two oranges.

This vector in the example is just a list of two numbers, so if we want to, we can represent it with a point in the plane with the corresponding coordinates, like so:



In the plane, we can imagine each point as some combination of apples and oranges (or parts thereof, for the points that don't have integer coordinates). Then each point corresponds to some vector. The collection of all such vectors—all the points in our apple-orange plane—is an example of a *vector space.*

Example There are 27 pieces of fruit in a barrel, and twice as many oranges as apples. How many apples and oranges are in the barrel?

How to solve this conundrum? We can re-write the question mathematically as follows:

$$\begin{aligned} x + y &= 27 \\ y &= 2x \end{aligned}$$

This is an example of a *Linear System*. It's a collection of equations in which variables are multiplied by constants and summed, and no variables are multiplied together: There are no powers of x or y greater than one, no fractional or negative powers of x or y, and no places where x and y are multiplied together.

Reading homework: problem 1.1

Notice that we can solve the system by manipulating the equations involved. First, notice that the second equation is the same as $-2x + y = 0$. Then if you subtract the second equation from the first, you get on the left side $x + y - (-2x + y) = 3x$, and on the left side you get $27 - 0 = 27$. Then $3x = 27$, so we learn that $x = 9$. Using the second equation, we then see that $y = 18$. Then there are 9 apples and 18 oranges.

Let's do it again, by working with the list of equations as an object in itself. First we rewrite the equations tidily:

$$\begin{aligned} x + y &= 27 \\ 2x - y &= 0 \end{aligned}$$

We can express this set of equations with a matrix as follows:

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 27 \\ 0 \end{pmatrix}$$

The square list of numbers is an example of a *matrix*. We can multiply the matrix by the vector to get back the linear system using the following rule for multiplying matrices by vectors:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix} \tag{1}$$

Reading homework: problem 1.2

The matrix is an example of a *Linear Transformation*, because it takes one vector and turns it into another in a "linear" way.

Our next task is to solve linear systems. We'll learn a general method called Gaussian Elimination.

References

Hefferon, Chapter One, Section 1
Beezer, Chapter SLE, Sections WILA and SSLE
Wikipedia, Systems of Linear Equations

Review Problems

1. Let M be a matrix and u and v vectors:
   $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, v = \begin{pmatrix} x \\ y \end{pmatrix}, u = \begin{pmatrix} w \\ z \end{pmatrix}.$

   (a) *Propose* a definition for $u + v$.

   (b) *Check* that your definition obeys $Mv + Mu = M(u + v)$.

2. *Matrix Multiplication:* Let M and N be matrices

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ and } N = \begin{pmatrix} e & f \\ g & h \end{pmatrix},$$

   and v a vector

$$v = \begin{pmatrix} x \\ c \end{pmatrix}.$$

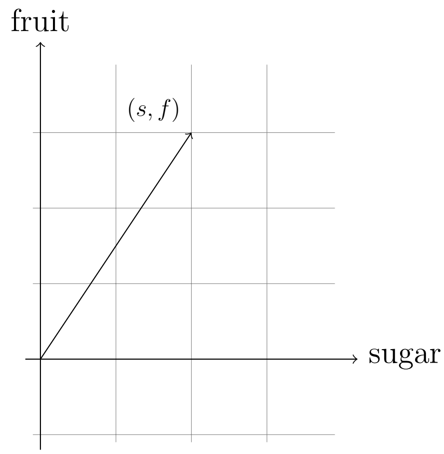
   Compute the vector Nv using the rule given above. Now multiply this vector by the matrix M, *i.e.*, compute the vector $M(Nv)$.

   Now recall that multiplication of ordinary numbers is associative, namely the order of brackets does not matter: $(xy)z = x(yz)$. Let us try to demand the same property for matrices and vectors, that is

$$M(Nv) = (MN)v\,.$$

   We need to be careful reading this equation because Nv is a vector and so is $M(Nv)$. Therefore the right hand side, $(MN)v$ should also be a vector. This means that MN must be a matrix; in fact it is the matrix obtained by multiplying the matrices M and N. Use your result for $M(Nv)$ to find the matrix MN.

3. Pablo is a nutritionist who knows that oranges always have twice as much sugar as apples. When considering the sugar intake of schoolchildren eating a barrel of fruit, he represents the barrel like so:



Find a linear transformation relating Pablo's representation to the one in the lecture. Write your answer as a matrix.

Hint: Let λ represent the amount of sugar in each apple.

4. There are methods for solving linear systems other than Gauss' method. One often taught in high school is to solve one of the equations for a variable, then substitute the resulting expression into other equations. That step is repeated until there is an equation with only one variable. From that, the first number in the solution is derived, and then back-substitution can be done. This method takes longer than Gauss' method, since it involves more arithmetic operations, and is also more likely to lead to errors. To illustrate how it can lead to wrong conclusions, we will use the system

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

   (a) Solve the first equation for x and substitute that expression into the second equation. Find the resulting y.

(b) Again solve the first equation for x, but this time substitute that expression into the third equation. Find this y.

What extra step must a user of this method take to avoid erroneously concluding a system has a solution?

2 Gaussian Elimination

2.1 Notation for Linear Systems

In Lecture 1 we studied the linear system

$$\begin{aligned} x + y &= 27 \\ 2x - y &= 0 \end{aligned}$$

and found that

$$\begin{aligned} x \quad &= 9 \\ y &= 18 \end{aligned}$$

We learned to write the linear system using a matrix and two vectors like so:

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 27 \\ 0 \end{pmatrix}$$

Likewise, we can write the solution as:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 9 \\ 18 \end{pmatrix}$$

The matrix $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is called the *Identity Matrix*. You can check that if v is any vector, then $Iv = v$.

A useful shorthand for a linear system is an *Augmented Matrix*, which looks like this for the linear system we've been dealing with:

$$\left(\begin{array}{cc|c} 1 & 1 & 27 \\ 2 & -1 & 0 \end{array}\right)$$

We don't bother writing the vector $\begin{pmatrix} x \\ y \end{pmatrix}$, since it will show up in any linear system we deal with. The solution to the linear system looks like this:

$$\left(\begin{array}{cc|c} 1 & 0 & 9 \\ 0 & 1 & 18 \end{array}\right)$$

Here's another example of an augmented matrix, for a linear system with three equations and four unknowns:

$$\left(\begin{array}{cccc|c} 1 & 3 & 2 & 0 & 9 \\ 6 & 2 & 0 & -2 & 0 \\ -1 & 0 & 1 & 1 & 3 \end{array}\right)$$

And finally, here's the general case. The number of equations in the linear system is the number of rows r in the augmented matrix, and the number of columns k in the matrix left of the vertical line is the number of unknowns.

$$\left(\begin{array}{cccc|c} a_1^1 & a_2^1 & \cdots & a_k^1 & b^1 \\ a_1^2 & a_2^2 & \cdots & a_k^2 & b^2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_1^r & a_2^r & \cdots & a_k^r & b^r \end{array}\right)$$

Reading homework: problem 2.1

Here's the idea: Gaussian Elimination is a set of rules for taking a general augmented matrix and turning it into a very simple augmented matrix consisting of the identity matrix on the left and a bunch of numbers (the solution) on the right.

Equivalence Relations for Linear Systems

It often happens that two mathematical objects will appear to be different but in fact are exactly the same. The best-known example of this are fractions. For example, the fractions $\frac{1}{2}$ and $\frac{6}{12}$ describe the same number. We could certainly call the two fractions *equivalent*.

In our running example, we've noticed that the two augmented matrices

$$\left(\begin{array}{cc|c} 1 & 1 & 27 \\ 2 & -1 & 0 \end{array}\right), \qquad \left(\begin{array}{cc|c} 1 & 0 & 9 \\ 0 & 1 & 18 \end{array}\right)$$

both contain the same information: $x = 9, y = 18$.

Two augmented matrices corresponding to linear systems that actually have solutions are said to be (row) *equivalent* if they have the *same* solutions. To denote this, we write:

$$\left(\begin{array}{cc|c} 1 & 1 & 27 \\ 2 & -1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 9 \\ 0 & 1 & 18 \end{array}\right)$$

The symbol $\sim$ is read "is equivalent to".

A small excursion into the philosophy of mathematical notation: Suppose I have a large pile of equivalent fractions, such as $\frac{2}{4}$, $\frac{27}{54}$, $\frac{100}{200}$, and so on. Most people will agree that their favorite way to write the number represented by all these different factors is $\frac{1}{2}$, in which the numerator and denominator are relatively prime. We usually call this a *reduced fraction*. This is an example of a *canonical form*, which is an extremely impressive way of saying "favorite way of writing it down". There's a theorem telling us that every rational number can be specified by a unique fraction whose numerator and denominator are relatively prime. To say that again, but slower, *every* rational number *has* a reduced fraction, and furthermore, that reduced fraction is *unique*.

2.2 Reduced Row Echelon Form

Since there are many different augmented matrices that have the same set of solutions, we should find a canonical form for writing our augmented matrices. This canonical form is called *Reduced Row Echelon Form*, or RREF for short. RREF looks like this in general:

$$\left(\begin{array}{ccccccc|c}
1 & * & 0 & * & 0 & \cdots & 0 & b^1 \\
0 & & 1 & * & 0 & \cdots & 0 & b^2 \\
0 & & 0 & & 1 & \cdots & 0 & b^3 \\
\vdots & & \vdots & & \vdots & & 0 & \vdots \\
 & & & & & & 1 & b^k \\
0 & & 0 & & 0 & \cdots & 0 & 0 \\
\vdots & & \vdots & & \vdots & & \vdots & \vdots \\
0 & & 0 & & 0 & \cdots & 0 & 0
\end{array}\right)$$

The first non-zero entry in each row is called the *pivot*. The asterisks denote arbitrary content which could be several columns long. The following properties describe the RREF.

1. In RREF, the pivot of any row is always 1.

2. The pivot of any given row is always to the right of the pivot of the row above it.

3. The pivot is the only non-zero entry in its column.

Example $\left(\begin{array}{ccc|c} 1 & 0 & 7 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right)$

Here is a NON-Example, which breaks all three of the rules:

$$\left(\begin{array}{ccc|c} 1 & 0 & 3 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

The RREF is a very useful way to write linear systems: it makes it very easy to write down the solutions to the system.

Example

$$\left(\begin{array}{cccc|c} 1 & 0 & 7 & 0 & 4 \\ 0 & 1 & 3 & 0 & 1 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

When we write this augmented matrix as a system of linear equations, we get the following:

$$\begin{array}{rcl} x \quad + 7z \quad\quad & = & 4 \\ y + 3z \quad\quad & = & 1 \\ w & = & 2 \end{array}$$

Solving from the bottom variables up, we see that $w = 2$ immediately. z is not a pivot, so it is still undetermined. Set $z = \lambda$. Then $y = 1 - 3\lambda$ and $x = 4 - 7\lambda$. More concisely:

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 0 \\ 2 \end{pmatrix} + \lambda \begin{pmatrix} -7 \\ -3 \\ 1 \\ 0 \end{pmatrix}$$

So we can read off the solution *set* directly from the RREF. (Notice that we use the word "set" because there is not just one solution, but one for every choice of λ.)

Reading homework: problem 2.2

You need to become very adept at reading off solutions of linear systems from the RREF of their augmented matrix. The general method is to work from the bottom up and set any non-pivot variables to unknowns. Here is another example.

Example

$$\left(\begin{array}{ccccc|c} 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Here we were not told the names of the variables, so lets just call them x_1, x_2, x_3, x_4, x_5. (There are always as many of these as there are columns in the matrix before the vertical line; the number of rows, on the other hand is the number of linear equations.)

To begin with we immediately notice that there are no pivots in the second and fourth columns so x_2 and x_4 are undetermined and we set them to

$$x_2 = \lambda_1 , \qquad x_4 = \lambda_2 .$$

(Note that you get to be creative here, we could have used λ and μ or any other names we like for a pair of unknowns.)

Working from the bottom up we see that the last row just says $0 = 0$, a well known fact! *Note that a row of zeros save for a non-zero entry after the vertical line would be mathematically inconsistent and indicates that the system has NO solutions at all.*

Next we see from the second last row that $x_5 = 3$. The second row says $x_3 = 2 - 2x_4 = 2 - 2\lambda_2$. The top row then gives $x_1 = 1 - x_2 - x_4 = 1 - \lambda_1 - \lambda_2$. Again we can write this solution as a vector

$$\begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 3 \end{pmatrix} + \lambda_1 \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 0 \\ -2 \\ 1 \\ 0 \end{pmatrix} .$$

Observe, that since no variables were given at the beginning, we do not really need to state them in our solution. As a challenge, look carefully at this solution and make sure you can see how every part of it comes from the original augmented matrix without every having to reintroduce variables and equations.

Perhaps unsurprisingly in light of the previous discussions of RREF, we have a theorem:

Theorem 2.1. *Every augmented matrix is row-equivalent to a* unique *augmented matrix in reduced row echelon form.*

Next lecture, we will prove it.

References

Hefferon, Chapter One, Section 1
Beezer, Chapter SLE, Section RREF
Wikipedia, Row Echelon Form

Review Problems

1. State whether the following augmented matrices are in RREF and compute their solution sets.

$$\left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 3 & 1 \\ 0 & 1 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 & 2 & 0 \end{array}\right),$$

$$\left(\begin{array}{cccccc|c} 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right),$$

$$\left(\begin{array}{ccccccc|c} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 & 0 & 2 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right).$$

2. Show that this pair of augmented matrices are row equivalent, assuming $ad - bc \neq 0$:

$$\left(\begin{array}{cc|c} a & b & e \\ c & d & f \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & \frac{de-bf}{ad-bc} \\ 0 & 1 & \frac{af-ce}{ad-bc} \end{array}\right)$$

3. Consider the augmented matrix: $\left(\begin{array}{cc|c} 2 & -1 & 3 \\ -6 & 3 & 1 \end{array}\right)$

   Give a *geometric* reason why the associated system of equations has no solution. (Hint, plot the three vectors given by the columns of this augmented matrix in the plane.) Given a general augmented matrix

$$\left(\begin{array}{cc|c} a & b & e \\ c & d & f \end{array}\right),$$

   can you find a condition on the numbers a, b, c and d that create the geometric condition you found?

4. List as many operations on augmented matrices that *preserve* row equivalence as you can. Explain your answers. Give examples of operations that break row equivalence.

5. Row equivalence of matrices is an example of an *equivalence relation*. Recall that a relation $\sim$ on a set of objects U is an equivalence relation if the following three properties are satisfied:

   - Reflexive: For any $x \in U$, we have $x \sim x$.
   - Symmetric: For any $x, y \in U$, if $x \sim y$ then $y \sim x$.
   - Transitive: For any x, y and $z \in U$, if $x \sim y$ and $y \sim z$ then $x \sim z$.

   (For a fuller discussion of equivalence relations, see Homework 0, Problem 4)

   Show that row equivalence of augmented matrices is an equivalence relation.

3 Elementary Row Operations

Our goal is to begin with an arbitrary matrix and apply operations that respect row equivalence until we have a matrix in Reduced Row Echelon Form (RREF). The three elementary row operations are:

- (Row Swap) Exchange any two rows.
- (Scalar Multiplication) Multiply any row by a non-zero constant.
- (Row Sum) Add a multiple of one row to another row.

Why do these preserve the linear system in question? Swapping rows is just changing the order of the equations begin considered, which certainly should not alter the solutions. Scalar multiplication is just multiplying the equation by the same number on both sides, which does not change the solution(s) of the equation. Likewise, if two equations share a common solution, adding one to the other preserves the solution.

There is a very simple process for row-reducing a matrix, working column by column. This process is called *Gauss–Jordan elimination* or simply Gaussian elimination.

1. If all entries in a given column are zero, then the associated variable is undetermined; make a note of the undetermined variable(s) and then ignore all such columns.
2. Swap rows so that the first entry in the first column is non-zero.
3. Multiply the first row by λ so that the pivot is 1.
4. Add multiples of the first row to each other row so that the first entry of every other row is zero.
5. Now ignore the first row and first column and repeat steps 1-5 until the matrix is in RREF.

Reading homework: problem 3.1

Example

$$\begin{aligned} 3x_3 &= 9 \\ x_1 + 5x_2 - 2x_3 &= 2 \\ \tfrac{1}{3}x_1 + 2x_2 \qquad &= 3 \end{aligned}$$

First we write the system as an augmented matrix:

$$\left(\begin{array}{ccc|c} 0 & 0 & 3 & 9 \\ 1 & 5 & -2 & 2 \\ \frac{1}{3} & 2 & 0 & 3 \end{array}\right) \underset{\sim}{R_1 \leftrightarrow R_3} \left(\begin{array}{ccc|c} \frac{1}{3} & 2 & 0 & 3 \\ 1 & 5 & -2 & 2 \\ 0 & 0 & 3 & 9 \end{array}\right)$$

$$\underset{\sim}{3R_1} \left(\begin{array}{ccc|c} 1 & 6 & 0 & 9 \\ 1 & 5 & -2 & 2 \\ 0 & 0 & 3 & 9 \end{array}\right)$$

$$\underset{\sim}{R_2 = R_2 - R_1} \left(\begin{array}{ccc|c} 1 & 6 & 0 & 9 \\ 0 & -1 & -2 & -7 \\ 0 & 0 & 3 & 9 \end{array}\right)$$

$$\underset{\sim}{-R_2} \left(\begin{array}{ccc|c} 1 & 6 & 0 & 9 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 3 & 9 \end{array}\right)$$

$$\underset{\sim}{R_1 = R_1 - 6R_2} \left(\begin{array}{ccc|c} 1 & 0 & -12 & -33 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 3 & 9 \end{array}\right)$$

$$\underset{\sim}{\tfrac{1}{3}R_3} \left(\begin{array}{ccc|c} 1 & 0 & -12 & -33 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

$$\underset{\sim}{R_1 = R_1 + 12R_3} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

$$\underset{\sim}{R_2 = R_2 - 2R_3} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

Now we're in RREF and can see that the solution to the system is given by $x_1 = 1$, $x_2 = 3$, and $x_3 = 1$; it happens to be a unique solution. Notice that we kept track of the steps we were taking; this is important for checking your work!

Example

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

$$
\overset{R_2-R_1;R_4-5R_2}{\sim} \left(\begin{array}{cccc|c} 1 & 0 & -1 & 2 & -1 \\ 0 & 1 & 2 & -3 & 3 \\ 0 & -1 & -2 & 3 & -3 \\ 0 & 2 & 4 & -6 & 6 \end{array}\right)
$$

$$
\overset{R_3+R_2;R_4-2R_3}{\sim} \left(\begin{array}{cccc|c} 1 & 0 & -1 & 2 & -1 \\ 0 & 1 & 2 & -3 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)
$$

Here the variables x_3 and x_4 are undetermined; the solution is not unique. Set $x_3 = \lambda$ and $x_4 = \mu$ where λ and μ are *arbitrary* real numbers. Then we can write x_1 and x_2 in terms of λ and μ as follows:

$$
\begin{aligned}
x_1 &= \lambda - 2\mu - 1 \\
x_2 &= -2\lambda + 3\mu + 3
\end{aligned}
$$

We can write the solution set with vectors like so:

$$
\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} -2 \\ 3 \\ 0 \\ 1 \end{pmatrix}
$$

This is (almost) our preferred form for writing the set of solutions for a linear system with many solutions.

Uniqueness of Gauss-Jordan Elimination

Theorem 3.1. *Gauss-Jordan Elimination produces a unique augmented matrix in RREF.*

Proof. Suppose Alice and Bob compute the RREF for a linear system but get different results, A and B. Working from the left, discard all columns except for the pivots and the first column in which A and B differ. By Review Problem 1b, removing columns does not affect row equivalence. Call the new, smaller, matrices $\hat{A}$ and $\hat{B}$. The new matrices should look this:

$$\hat{A} = \left(\begin{array}{c|c} I_N & a \\ 0 & 0 \end{array}\right) \text{ and } \hat{B} = \left(\begin{array}{c|c} I_N & b \\ 0 & 0 \end{array}\right),$$

where I_N is an $N \times N$ identity matrix and a and b are vectors.

Now if $\hat{A}$ and $\hat{B}$ have the same solution, then we must have $a = b$. But this is a contradiction! Then $A = B$. $\square$

References

Hefferon, Chapter One, Section 1.1 and 1.2
Beezer, Chapter SLE, Section RREF
Wikipedia, Row Echelon Form
Wikipedia, Elementary Matrix Operations

Review Problems

1. (Row Equivalence)

   (a) Solve the following linear system using Gauss-Jordan elimination:

$$\begin{aligned} 2x_1 + 5x_2 - 8x_3 + 2x_4 + 2x_5 &= 0 \\ 6x_1 + 2x_2 - 10x_3 + 6x_4 + 8x_5 &= 6 \\ 3x_1 + 6x_2 + 2x_3 + 3x_4 + 5x_5 &= 6 \\ 3x_1 + 1x_2 - 5x_3 + 3x_4 + 4x_5 &= 3 \\ 6x_1 + 7x_2 - 3x_3 + 6x_4 + 9x_5 &= 9 \end{aligned}$$

   Be sure to set your work out carefully with equivalence signs $\sim$ between each step, labeled by the row operations you performed.

   (b) Check that the following two matrices are row-equivalent:

$$\left(\begin{array}{ccc|c} 1 & 4 & 7 & 10 \\ 2 & 9 & 6 & 0 \end{array}\right) \text{ and } \left(\begin{array}{ccc|c} 0 & -1 & 8 & 20 \\ 4 & 18 & 12 & 0 \end{array}\right)$$

Now remove the third column from each matrix, and show that the resulting two matrices (shown below) are row-equivalent:

$$\left(\begin{array}{cc|c} 1 & 4 & 10 \\ 2 & 9 & 0 \end{array}\right) \text{ and } \left(\begin{array}{cc|c} 0 & -1 & 20 \\ 4 & 18 & 0 \end{array}\right)$$

Now remove the fourth column from each of the original two matrices, and show that the resulting two matrices, viewed as augmented matrices (shown below) are row-equivalent:

$$\left(\begin{array}{cc|c} 1 & 4 & 7 \\ 2 & 9 & 6 \end{array}\right) \text{ and } \left(\begin{array}{cc|c} 0 & -1 & 8 \\ 4 & 18 & 12 \end{array}\right)$$

Explain why row-equivalence is never affected by removing columns.

(c) Check that the matrix $\left(\begin{array}{cc|c} 1 & 4 & 10 \\ 3 & 13 & 9 \\ 4 & 17 & 20 \end{array}\right)$ has no solutions. If you remove one of the rows of this matrix, does the new matrix have any solutions? In general, can row equivalence be affected by removing rows? Explain why or why not.

2. (Gaussian Elimination) Another method for solving linear systems is to use row operations to bring the augmented matrix to row echelon form. In row echelon form, the pivots are not necessarily set to one, and we only require that all entries left of the pivots are zero, not necessarily entries above a pivot. Provide a counterexample to show that row echelon form is not unique.

   Once a system is in row echelon form, it can be solved by "back substitution." Write the following row echelon matrix as a system of equations, then solve the system using back-substitution.

$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 6 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 3 & 3 \end{array}\right)$$

3. Explain why the linear system has no solutions:

$$\left(\begin{array}{ccc|c} 1 & 0 & 3 & 1 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 6 \end{array}\right)$$

For which values of k does the system below have a solution?

$$\begin{array}{rcrcrcr} x & - & 3y & & & = & 6 \\ x & & & + & 3z & = & -3 \\ 2x & + & ky & + & (3-k)z & = & 1 \end{array}$$

4 Solution Sets for Systems of Linear Equations

For a system of equations with r equations and k unknowns, one can have a number of different outcomes. For example, consider the case of r equations in three variables. Each of these equations is the equation of a plane in three-dimensional space. To find solutions to the system of equations, we look for the common intersection of the planes (if an intersection exists). Here we have five different possibilities:

1. **No solutions.** Some of the equations are contradictory, so no solutions exist.

2. **Unique Solution.** The planes have a unique point of intersection.

3. **Line.** The planes intersect in a common line; any point on that line then gives a solution to the system of equations.

4. **Plane.** Perhaps you only had one equation to begin with, or else all of the equations coincide geometrically. In this case, you have a plane of solutions, with two free parameters.

5. **All of $\mathbb{R}^3$.** If you start with no information, then any point in $\mathbb{R}^3$ is a solution. There are three free parameters.

In general, for systems of equations with k unknowns, there are $k + 2$ possible outcomes, corresponding to the number of free parameters in the solutions set, plus the possibility of no solutions. These types of "solution sets" are hard to visualize, but luckily "hyperplanes" behave like planes in $\mathbb{R}^3$ in many ways.

Reading homework: problem 4.1

4.1 Non-Leading Variables

Variables that are not a pivot in the reduced row echelon form of a linear system are *free*. We set them equal to arbitrary parameters $\mu_1, \mu_2, \ldots$

Example $\left(\begin{array}{cccc|c}1 & 0 & 1 & -1 & 1\\ 0 & 1 & -1 & 1 & 1\\ 0 & 0 & 0 & 0 & 0\end{array}\right)$

Here, x_1 and x_2 are the pivot variables and x_3 and x_4 are non-leading variables, and thus free. The solutions are then of the form $x_3 = \mu_1$, $x_4 = \mu_2$, $x_2 = 1+\mu_1-\mu_2$, $x_1 = 1 - \mu_1 + \mu_2$.

The preferred way to write a solution set is with set notation. Let S be the set of solutions to the system. Then:

$$S = \left\{ \begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix} = \begin{pmatrix}1\\1\\0\\0\end{pmatrix} + \mu_1 \begin{pmatrix}-1\\1\\1\\0\end{pmatrix} + \mu_2 \begin{pmatrix}1\\-1\\0\\1\end{pmatrix} \right\}$$

We have already seen how to write a linear system of two equations in two unknowns as a matrix multiplying a vector. We can apply exactly the same idea for the above system of three equations in four unknowns by calling

$$M = \begin{pmatrix}1 & 0 & 1 & -1\\ 0 & 1 & -1 & 1\\ 0 & 0 & 0 & 0\end{pmatrix}, \quad X = \begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix} \text{ and } V = \begin{pmatrix}1\\1\\0\end{pmatrix}.$$

Then if we take for the product of the matrix M with the vector X of unknowns

$$MX = \begin{pmatrix}1 & 0 & 1 & -1\\ 0 & 1 & -1 & 1\\ 0 & 0 & 0 & 0\end{pmatrix} \begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix} = \begin{pmatrix}x_1 + x_3 - x_4\\ x_2 - x_3 + x_4\\ 0\end{pmatrix}$$

our system becomes simply

$$MX = V\,.$$

Stare carefully at our answer for the product MX above. First you should notice that each of the three rows corresponds to the left hand side of one of the equations in the system. Also observe that each entry was obtained by matching the entries in the corresponding row of M with the column entries of X. For example, using the second row of M we obtained the second entry of MX

$$0 \;\; 1 \;\; -1 \;\; 1 \quad \begin{matrix}x_1\\x_2\\x_3\\x_4\end{matrix} \quad \longmapsto \; x_2 - x_3 + x_4\,.$$

In Lecture 8 we will study matrix multiplication in detail, but you can already try to discover the main rules for your for yourself by working through Review Question 3 on multiplying matrices by vectors.

Given two vectors we can *add* them term-by-term:

$$\begin{pmatrix} a^1 \\ a^2 \\ a^3 \\ \vdots \\ a^r \end{pmatrix} + \begin{pmatrix} b^1 \\ b^2 \\ b^3 \\ \vdots \\ b^r \end{pmatrix} = \begin{pmatrix} a^1 + b^1 \\ a^2 + b^2 \\ a^3 + b^3 \\ \vdots \\ a^r + b^r \end{pmatrix}$$

We can also multiply a vector by a scalar, like so:

$$\lambda \begin{pmatrix} a^1 \\ a^2 \\ a^3 \\ \vdots \\ a^r \end{pmatrix} = \begin{pmatrix} \lambda a^1 \\ \lambda a^2 \\ \lambda a^3 \\ \vdots \\ \lambda a^r \end{pmatrix}$$

Then yet another way to write the solution set for the example is:

$$X = X_0 + \mu_1 Y_1 + \mu_2 Y_2$$

where

$$X_0 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, Y_1 = \begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, Y_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

Definition Let X and Y be vectors and α and β be scalars. A function f is *linear* if

$$f(\alpha X + \beta Y) = \alpha f(X) + \beta f(Y)$$

This is called the *linearity property* for matrix multiplication.

The notion of linearity is a core concept in this course. Make sure you understand what it means and how to use it in computations!

Example Consider our example system above with

$$M = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \text{ and } Y = \begin{pmatrix} y_1 \\ y_2 \\ y^3 \\ y^4 \end{pmatrix},$$

and take for the function of vectors

$$f(X) = MX.$$

Now let us check the linearity property for f. The property needs to hold for *any* scalars α and β, so for simplicity let us concentrate first on the case $\alpha = \beta = 1$. This means that we need to compare the following two calculations:

1. First add $X + Y$, then compute $f(X + Y)$.
2. First compute $f(X)$ and $f(Y)$, then compute the sum $f(X) + f(Y)$.

The second computation is slightly easier:

$$f(X) = MX = \begin{pmatrix} x_1 + x_3 - x_4 \\ x_2 - x_3 + x_4 \\ 0 \end{pmatrix} \text{ and } f(Y) = MY = \begin{pmatrix} y_1 + y_3 - y_4 \\ y_2 - y_3 + y_4 \\ 0 \end{pmatrix},$$

(using our result above). Adding these gives

$$f(X) + f(Y) = \begin{pmatrix} x_1 + x_3 - x_4 + y_1 + y_3 - y_4 \\ x_2 - x_3 + x_4 + y_2 - y_3 + y_4 \\ 0 \end{pmatrix}.$$

Next we perform the first computation beginning with:

$$X + Y = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \\ x_4 + y_4 \end{pmatrix},$$

from which we calculate

$$f(X + Y) = \begin{pmatrix} x_1 + y_2 + x_3 + y_3 - (x_4 + y_4) \\ x_2 + y_2 - (x_3 + y_3) + x_4 + y_4 \\ 0 \end{pmatrix}.$$

Distributing the minus signs and remembering that the order of adding numbers like $x_1, x_2, \ldots$ does not matter, we see that the two computations give exactly the same answer.

Of course, you should complain that we took a special choice of α and β. Actually, to take care of this we only need to check that $f(\alpha X) = \alpha f(X)$. It is your job to explain this in Review Question 1

Later we will show that matrix multiplication is always linear. Then we will know that:

$$M(\alpha X + \beta Y) = \alpha MX + \beta MY$$

Then the two equations $MX = V$ and $X = X_0 + \mu_1 Y_1 + \mu_2 Y_2$ together say that:

$$MX_0 + \mu_1 MY_1 + \mu_2 MY_2 = V$$

for *any* $\mu_1, \mu_2 \in \mathbb{R}$. Choosing $\mu_1 = \mu_2 = 0$, we obtain

$$MX_0 = V\,.$$

Here, X_0 is an example of what is called a *particular solution* to the system.

Given the particular solution to the system, we can then deduce that $\mu_1 MY_1 + \mu_2 MY_2 = 0$. Setting $\mu_1 = 0, \mu_2 = 1$, and recalling the particular solution $MX_0 = V$, we obtain

$$MY_1 = 0\,.$$

Likewise, setting $\mu_1 = 1, \mu_2 = 0$, we obtain

$$MY_2 = 0\,.$$

Here Y_1 and Y_2 are examples of what are called *homogeneous* solutions to the system. They *do not* solve the original equation $MX = V$, but instead its associated homogeneous system of equations $MY = 0$.

Example Consider the linear system with the augmented matrix we've been working with.

$$\begin{array}{rrrrl} x & & +z & -w & = 1 \\ & y & -z & +w & = 1 \end{array}$$

Recall that the system has the following solution set:

$$S = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \mu_1 \begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \end{pmatrix} + \mu_2 \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Then $MX_0 = V$ says that $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ solves the original system of equations, which is certainly true, but this is not the only solution.

$MY_1 = 0$ says that $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$ solves the homogeneous system.

$MY_2 = 0$ says that $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$ solves the homogeneous system.

Notice how adding any multiple of a homogeneous solution to the particular solution yields another particular solution.

Definition Let M a matrix and V a vector. Given the linear system $MX = V$, we call X_0 a *particular solution* if $MX_0 = V$. We call Y a *homogeneous solution* if $MY = 0$. The linear system

$$MX = 0$$

is called the (associated) *homogeneous system*.

If X_0 is a particular solution, then the general solution to the system is[1]:

$$S = \{X_0 + Y : MY = 0\}$$

In other words, the general solution = particular + homogeneous.

Reading homework: problem 4.2

[1]The notation $S = \{X_0 + Y : MY = 0\}$ is read, "S is the set of all $X_0 + Y$ such that $MY = 0$," and means exactly that. Sometimes a pipe $|$ is used instead of a colon.

References

Hefferon, Chapter One, Section I.2
Beezer, Chapter SLE, Section TSS
Wikipedia, Systems of Linear Equations

Review Questions

1. Let $f(X) = MX$ where

$$M = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ and } X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} .$$

   Suppose that α is any number. Compute the following four quantities:

$$\alpha X\,,\ f(X)\,,\ \alpha f(X) \text{ and } f(\alpha X)\,.$$

   Check your work by verifying that

$$\alpha f(X) = f(\alpha X)\,.$$

   Now explain why the result checked in the Lecture, namely

$$f(X+Y) = f(X) + f(Y)\,,$$

   and your result $f(\alpha X) = \alpha f(X)$ together imply

$$f(\alpha X + \beta Y) = \alpha f(X) + \beta f(Y)\,.$$

2. Write down examples of augmented matrices corresponding to each of the five types of solution sets for systems of equations with three unknowns.

3. Let

$$M = \begin{pmatrix} a_1^1 & a_2^1 & \cdots & a_k^1 \\ a_1^2 & a_2^2 & \cdots & a_k^2 \\ \vdots & \vdots & & \vdots \\ a_1^r & a_2^r & \cdots & a_k^r \end{pmatrix}, \qquad X = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^k \end{pmatrix}$$

Propose a rule for MX so that $MX = 0$ is equivalent to the linear system:

$$\begin{array}{cccc} a^1_1x^1 & +a^1_2x^2 & \cdots +a^1_kx^k & = 0 \\ a^2_1x^1 & +a^2_2x^2 & \cdots +a^2_kx^k & = 0 \\ \vdots & \vdots & \vdots & \vdots \\ a^r_1x^1 & +a^r_2x^2 & \cdots +a^r_kx^k & = 0 \end{array}$$

Show that your rule for multiplying a matrix by a vector obeys the linearity property.

Note that in this problem, x^2 does not denote the square of x. Instead x^1, x^2, x^3, etc... denote different variables. Although confusing at first, this notation was invented by Albert Einstein who noticed that quantities like $a^2_1x^1 + a^2_2x^2 \cdots + a^2_kx^k$ could be written in summation notation as $\sum_{j=1}^k a^2_jx^j$. Here j is called a summation index. Einstein observed that you could even drop the summation sign $\sum$ and simply write $a^2_jx^j$.

4. Use the rule you developed in the problem 3 to compute the following products

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 14 \\ 14 \\ 21 \\ 35 \\ 62 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 42 & 97 & 2 & -23 & 46 \\ 0 & 1 & 3 & 1 & 0 & 33 \\ 11 & \pi & 1 & 0 & 46 & 29 \\ -98 & 12 & 0 & 33 & 99 & 98 \\ \log 2 & 0 & \sqrt{2} & 0 & e & 23 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 7 & 8 & 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 & 17 & 18 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Now that you are good at multiplying a matrix with a column vector, try your hand at a product of two matrices

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 7 & 8 & 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 & 17 & 18 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Hint, to do this problem, either skip ahead and read Lecture 8, or just view the matrix on the right as three column vectors next to one another.

5. The *standard basis vector* e_i is a column vector with a one in the ith row, and zeroes everywhere else. Using the rule for multiplying a matrix times a vector in problem 3, find a simple rule for multiplying Me_i, where M is the general matrix defined there.

5 Vectors in Space, n-Vectors

In vector calculus classes, you encountered three-dimensional vectors. Now we will develop the notion of n-vectors and learn some of their properties.

We begin by looking at the space $\mathbb{R}^n$, which we can think of as the space of points with n coordinates. We then specify an *origin* O, a favorite point in $\mathbb{R}^n$. Now given any other point P, we can draw a *vector* v from O to P. Just as in $\mathbb{R}^3$, a vector has a *magnitude* and a *direction*.

If O has coordinates $(o^1, \ldots, o^n)$ and p has coordinates $(p^1, \ldots, p^n)$, then the *components* of the vector v are $\begin{pmatrix} p^1 - o^1 \\ p^2 - o^2 \\ \vdots \\ p^n - o^n \end{pmatrix}$. This construction allows us to put the origin anywhere that seems most convenient in $\mathbb{R}^n$, not just at the point with zero coordinates.

Do not be confused by our use of a superscript to label components of a vector. Here v^2 denotes the second component of a vector v, rather than a number v squared!

Most importantly, we can *add* vectors and *multiply* vectors by a scalar:

Definition Given two vectors a and b whose components are given by

$$a = \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} \text{ and } b = \begin{pmatrix} b^1 \\ \vdots \\ b^n \end{pmatrix}$$

their *sum* is

$$a + b = \begin{pmatrix} a^1 + b^1 \\ \vdots \\ a^n + b^n \end{pmatrix}.$$

Given a scalar λ, the *scalar multiple*

$$\lambda a = \begin{pmatrix} \lambda a^1 \\ \vdots \\ \lambda a^n \end{pmatrix}.$$

Example Let

$$a = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \text{ and } b = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix}.$$

Then, for example

$$a + b = \begin{pmatrix} 5 \\ 5 \\ 5 \\ 5 \end{pmatrix} \text{ and } 3a - 2b = \begin{pmatrix} -5 \\ 0 \\ 5 \\ 10 \end{pmatrix}.$$

Notice that these are the same rules we saw in Lecture 4! In Lectures 1-4, we thought of a vector as being a list of numbers which captured information about a linear system. Now we are thinking of a vector as a magnitude and a direction in $\mathbb{R}^n$, and luckily the same rules apply.

A special vector is the *zero vector* connecting the origin to itself. All of its components are zero. Notice that with respect to the usual notions of Euclidean geometry, it is the only vector with zero magnitude, and the only one which points in no particular direction. Thus, any single vector determines a line, *except* the zero-vector. Any scalar multiple of a non-zero vector lies in the line determined by that vector.

The line determined by a non-zero vector v through a point P can be written as $\{P + tv | t \in \mathbb{R}\}$. For example, $\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \middle| t \in \mathbb{R} \right\}$ describes a line in 4-dimensional space parallel to the x-axis.

Given two non-zero vectors, they will *usually* determine a plane, unless both vectors are in the same line. In this case, one of the vectors can be realized as a scalar multiple of the other. The sum of u and v corresponds to laying the two vectors head-to-tail and drawing the connecting vector. If u and v determine a plane, then their sum lies in plane determined by u and v.

The plane determined by two vectors u and v can be written as

$$\{P + su + tv | s, t \in \mathbb{R}\}.$$

Example

$$\left\{ \begin{pmatrix} 3\\1\\4\\1\\5\\9 \end{pmatrix} + s \begin{pmatrix} 1\\0\\0\\0\\0\\0 \end{pmatrix} + t \begin{pmatrix} 0\\1\\0\\0\\0\\0 \end{pmatrix} \middle| s,t \in \mathbb{R} \right\}$$

describes a plane in 6-dimensional space parallel to the xy-plane.

We can generalize the notion of a plane:

Definition A set of k vectors $v_1, \ldots, v_k$ in $\mathbb{R}^n$ with $k \leq n$ determines a k-dimensional *hyperplane*, unless any of the vectors v_i lives in the same hyperplane determined by the other vectors. If the vectors do determine a k-dimensional hyperplane, then any point in the hyperplane can be written as:

$$\{P + \sum_{i=1}^{k} \lambda_i v_i | \lambda_i \in \mathbb{R}\}$$

5.1 Directions and Magnitudes

Consider the Euclidean length of a vector:

$$\|v\| = \sqrt{(v^1)^2 + (v^2)^2 + \cdots (v^n)^2} \;=\; \sqrt{\sum_{i=1}^{n} (v^i)^2}\,.$$

Using the Law of Cosines, we can then figure out the angle between two vectors. Given two vectors v and u that span a plane in $\mathbb{R}^n$, we can then connect the ends of v and u with the vector $v - u$. Then the Law of Cosines states that:

$$\|v - u\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\,\|v\| \cos\theta$$

Then isolate $\cos\theta$:

$$
\begin{aligned}
\|v-u\|^2 - \|u\|^2 - \|v\|^2 &= (v^1-u^1)^2 + \cdots + (v^n-u^n)^2 \\
&\qquad -((u^1)^2 + \cdots + (u^n)^2) \\
&\qquad -((v^1)^2 + \cdots + (v^n)^2) \\
&= -2u^1v^1 - \cdots - 2u^nv^n
\end{aligned}
$$

Thus,

$$\|u\|\,\|v\|\cos\theta = u^1v^1 + \cdots + u^nv^n\,.$$

Note that in the above discussion, we have assumed (correctly) that Euclidean lengths in $\mathbb{R}^n$ give the usual notion of lengths of vectors in the plane. This now motivates the definition of the dot product.

Definition The *dot product* of two vectors $u = \begin{pmatrix} u^1 \\ \vdots \\ u^n \end{pmatrix}$ and $v = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}$ is

$$u \cdot v = u^1v^1 + \cdots + u^nv^n\,.$$

The *length* of a vector

$$\|v\| = \sqrt{v \cdot v}\,.$$

The *angle* θ between two vectors is determined by the formula

$$u \cdot v = \|u\|\|v\|\cos\theta\,.$$

The dot product has some important properties:

1. The dot product is *symmetric*, so

$$u \cdot v = v \cdot u\,,$$

2. *Distributive* so

$$u \cdot (v+w) = u \cdot v + u \cdot w\,,$$

3. *Bilinear*, which is to say, linear in both u and v. Thus

$$u \cdot (cv + dw) = c\,u \cdot v + d\,u \cdot w\,,$$

and

$$(cu + dw) \cdot v = c\,u \cdot v + d\,w \cdot v\,.$$

4. *Positive Definite*:

$$u \cdot u \geq 0\,,$$

and $u \cdot u = 0$ only when u itself is the 0-vector.

There are, in fact, many different useful ways to define lengths of vectors. Notice in the definition above that we first defined the dot product, and then defined everything else in terms of the dot product. So if we change our idea of the dot product, we change our notion of length and angle as well. The dot product determines the *Euclidean length and angle* between two vectors.

Other definitions of length and angle arise from *inner products*, which have all of the properties listed above (except that in some contexts the positive definite requirement is relaxed). Instead of writing $\cdot$ for other inner products, we usually write $\langle u, v\rangle$ to avoid confusion.

Reading homework: problem 5.1

Example Consider a four-dimensional space, with a special direction which we will call "time". The *Lorentzian inner product* on $\mathbb{R}^4$ is given by $\langle u, v\rangle = u^1v^1 + u^2v^2 + u^3v^3 - u^4v^4$. This is of central importance in Einstein's theory of special relativity.

As a result, the "squared-length" of a vector with coordinates x, y, z and t is $\|v\|^2 = x^2 + y^2 + z^2 - t^2$. Notice that it is possible for $\|v\|^2 \leq 0$ for non-vanishing v!

Theorem 5.1 (Cauchy-Schwartz Inequality)**.** *For non-zero vectors u and v with an inner-product $\langle\ ,\ \rangle$,*

$$\frac{|\langle u, v\rangle|}{\|u\|\,\|v\|} \leq 1$$

Proof. The easiest proof would use the definition of the angle between two vectors and the fact that $\cos\theta \leq 1$. However, strictly speaking speaking we did not check our assumption that we could apply the Law of Cosines to the Euclidean length in $\mathbb{R}^n$. There is, however a simple algebraic proof. Let α be any real number and consider the following positive, quadratic polynomial in α

$$0 \leq \langle u + \alpha v, u + \alpha v\rangle = \langle u, u\rangle + 2\alpha\langle u, v\rangle + \alpha^2\langle v, v\rangle\,.$$

You should carefully check for yourself exactly which properties of an inner product were used to write down the above inequality!

Next, a tiny calculus computation shows that any quadratic $a\alpha^2 + 2b\alpha + c$ takes its minimal value $c - \frac{b^2}{a}$ when $\alpha = -\frac{b}{a}$. Applying this to the above quadratic gives

$$0 \leq \langle u, u \rangle - \frac{\langle u, v \rangle^2}{\langle v, v \rangle} .$$

Now it is easy to rearrange this inequality to reach the Cauchy–Schwartz one above. $\square$

Theorem 5.2 (Triangle Inequality)**.** *Given vectors u and v, we have:*

$$\|u + v\| \leq \|u\| + \|v\|$$

Proof.

$$\begin{aligned} \|u+v\|^2 &= (u+v) \cdot (u+v) \\ &= u \cdot u + 2u \cdot v + v \cdot v \\ &= \|u\|^2 + \|v\|^2 + 2\,\|u\|\,\|v\| \cos\theta \\ &= (\|u\| + \|v\|)^2 + 2\,\|u\|\,\|v\|(\cos\theta - 1) \\ &\leq (\|u\| + \|v\|)^2 \end{aligned}$$

Then the square of the left-hand side of the triangle inequality is $\leq$ the right-hand side, and both sides are positive, so the result is true. $\square$

Example Let

$$a = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \text{ and } b = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} ,$$

so that

$$a \cdot a = b \cdot b = 1 + 2^2 + 3^2 + 4^2 = 30$$
$$\Rightarrow \|a\| = \sqrt{30} = \|b\| \text{ and } \big(\|a\| + \|b\|\big)^2 = (2\sqrt{30})^2 = 120 .$$

Since

$$a + b = \begin{pmatrix} 5 \\ 5 \\ 5 \\ 5 \end{pmatrix} ,$$

we have

$$\|a+b\|^2 = 5^2 + 5^2 + 5^2 + 5^2 = 100 < 120 = \big(\|a\| + \|b\|\big)^2$$

as predicted by the triangle inequality.

Notice also that $a \cdot b = 1.4 + 2.3 + 3.2 + 4.1 = 20 < \sqrt{30}.\sqrt{30} = 30 = \|a\|\,\|b\|$ in accordance with the Cauchy–Schwartz inequality.

Reading homework: problem 5.2

References

Hefferon: Chapter One.II
Beezer: Chapter V, Section VO, Subsection VEASM
Beezer: Chapter V, Section O, Subsections IP-N
Relevant Wikipedia Articles:

- Dot Product
- Inner Product Space
- Minkowski Metric

Review Questions

1. When he was young, Captain Conundrum mowed lawns on weekends to help pay his college tuition bills. He charged his customers according to the size of their lawns at a rate of 5¢ per square foot and meticulously kept a record of the areas of their lawns in an ordered list:

$$A = (200, 300, 50, 50, 100, 100, 200, 500, 1000, 100)\,.$$

He also listed the number of times he mowed each lawn in a given year, for the year 1988 that ordered list was

$$f = (20, 1, 2, 4, 1, 5, 2, 1, 10, 6)\,.$$

(a) Pretend that A and f are vectors and compute $A \cdot f$.

(b) What quantity does the dot product $A \cdot f$ measure?

(c) How much did Captain Conundrum earn from mowing lawns in 1988? Write an expression for this amount in terms of the vectors A and f.

(d) Suppose Captain Conundrum charged different customers different rates. How could you modify the expression in part 1c to compute the Captain's earnings?

2. (2) Find the angle between the diagonal of the unit square in $\mathbb{R}^2$ and one of the coordinate axes.

   (3) Find the angle between the diagonal of the unit cube in $\mathbb{R}^3$ and one of the coordinate axes.

   (n) Find the angle between the diagonal of the unit (hyper)-cube in $\mathbb{R}^n$ and one of the coordinate axes.

   (∞) What is the limit as $n \to \infty$ of the angle between the diagonal of the unit (hyper)-cube in $\mathbb{R}^n$ and one of the coordinate axes?

3. Consider the matrix $M = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ and the vector $X = \begin{pmatrix} x \\ y \end{pmatrix}$.

   (a) Sketch X and MX in $\mathbb{R}^2$ for several values of X and θ.

   (b) Compute $\frac{||MX||}{||X||}$ for arbitrary values of X and θ.

   (c) Explain your result for (b) and describe the action of M geometrically.

4. Suppose in $\mathbb{R}^2$ I measure the x direction in inches and the y direction in miles. Approximately what is the real-world angle between the vectors $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$? What is the angle between these two vectors according to the dot-product? Give a definition for an inner product so that the angles produced by the inner product are the actual angles between vectors.

5. (Lorentzian Strangeness). For this problem, consider $\mathbb{R}^n$ with the Lorentzian inner product and metric defined above.

   (a) Find a non-zero vector in two-dimensional Lorentzian space-time with zero length.

   (b) Find and sketch the collection of all vectors in two-dimensional Lorentzian space-time with zero length.

   (c) Find and sketch the collection of all vectors in three-dimensional Lorentzian space-time with zero length.

6 Vector Spaces

Thus far we have thought of vectors as lists of numbers in $\mathbb{R}^n$. As it turns out, the notion of a vector applies to a much more general class of structures than this. The main idea is to define vectors based on their most important properties. Once complete, our new definition of vectors will include vectors in $\mathbb{R}^n$, but will also cover many other extremely useful notions of vectors. We do this in the hope of creating a mathematical structure applicable to a wide range of real-world problems.

The two key properties of vectors are that they can be added together and multiplied by scalars. So we make the following definition.

Definition A *vector space* (over $\mathbb{R}$) is a set V with two operations $+$ and $\cdot$ satisfying the following properties for all $u, v \in V$ and $c, d \in \mathbb{R}$:

(+i) (Additive Closure) $u + v \in V$. (Adding two vectors gives a vector.)

(+ii) (Additive Commutativity) $u + v = v + u$. (Order of addition doesn't matter.)

(+iii) (Additive Associativity) $(u + v) + w = u + (v + w)$ (Order of adding many vectors doesn't matter.)

(+iv) (Zero) There is a special vector $0_V \in V$ such that $u + 0_V = u$ for all u in V.

(+v) (Additive Inverse) For every $u \in V$ there exists $w \in V$ such that $u + w = 0_V$.

($\cdot$ i) (Multiplicative Closure) $c \cdot v \in V$. (Scalar times a vector is a vector.)

($\cdot$ ii) (Distributivity) $(c+d)\cdot v = c\cdot v+d\cdot v$. (Scalar multiplication distributes over addition of scalars.)

($\cdot$ iii) (Distributivity) $c\cdot(u+v) = c\cdot u+c\cdot v$. (Scalar multiplication distributes over addition of vectors.)

($\cdot$ iv) (Associativity) $(cd) \cdot v = c \cdot (d \cdot v)$.

($\cdot$ v) (Unity) $1 \cdot v = v$ for all $v \in V$.

Remark Don't confuse the scalar product $\cdot$ with the dot product $\bullet$. The scalar product is a function that takes a vector and a number and returns a vector. (In notation, this can be written $\cdot\colon \mathbb{R} \times V \to V$.) On the other hand, the dot product takes two vectors and returns a number. (In notation: $\bullet\colon V \times V \to \mathbb{R}$.)

Once the properties of a vector space have been verified, we'll just write scalar multiplication with juxtaposition $cv = c \cdot v$, though, to avoid confusing the notation.

Remark It isn't hard to devise strange rules for addition or scalar multiplication that break some or all of the rules listed above.

One can also find many interesting vector spaces, such as the following.

Example

$$V = \{f | f\colon \mathbb{N} \to \mathbb{R}\}$$

Here the vector space is the set of functions that take in a natural number n and return a real number. The addition is just addition of functions: $(f_1+f_2)(n) = f_1(n)+f_2(n)$. Scalar multiplication is just as simple: $c \cdot f(n) = cf(n)$.

We can think of these functions as infinite sequences: $f(0)$ is the first term, $f(1)$ is the second term, and so on. Then for example the function $f(n) = n^3$ would look like this:

$$f = \{0, 1, 8, 27, \ldots, n^3, \ldots\}.$$

Thinking this way, V is the space of all infinite sequences.

Let's check some axioms.

(+i) (Additive Closure) $f_1(n) + f_2(n)$ is indeed a function $\mathbb{N} \to \mathbb{R}$, since the sum of two real numbers is a real number.

(+iv) (Zero) We need to propose a zero vector. The constant zero function $g(n) = 0$ works because then $f(n) + g(n) = f(n) + 0 = f(n)$.

The other axioms that should be checked come down to properties of the real numbers.

Reading homework: problem 6.1

Example Another very important example of a vector space is the space of all differentiable functions:

$$\left\{ f | f\colon \mathbb{R} \to \mathbb{R},\ \frac{d}{dx} f \text{ exists} \right\}.$$

The addition is point-wise

$$(f+g)(x) = f(x) + g(x)\,,$$

as is scalar multiplication

$$c \cdot f(x) = cf(x)\,.$$

From calculus, we know that the sum of any two differentiable functions is differentiable, since the derivative distributes over addition. A scalar multiple of a function is also differentiable, since the derivative commutes with scalar multiplication ($\frac{d}{dx}(cf) = c\frac{d}{dx}f$). The zero function is just the function such that $0(x) = 0$ for every x. The rest of the vector space properties are inherited from addition and scalar multiplication in $\mathbb{R}$.

In fact, the set of functions with at least k derivatives is always a vector space, as is the space of functions with infinitely many derivatives.

Vector Spaces Over Other Fields Above, we defined vector spaces over the real numbers. One can actually define vector spaces over any *field*. A field is a collection of "numbers" satisfying a number of properties.

One other example of a field is the complex numbers,

$$\mathbb{C} = \left\{ x + iy | i^2 = -1, x, y \in \mathbb{R} \right\}.$$

In quantum physics, vector spaces over $\mathbb{C}$ describe all possible states a system of particles can have.

For example,

$$V = \left\{ \begin{pmatrix} \lambda \\ \mu \end{pmatrix} : \lambda, \mu \in \mathbb{C} \right\}$$

describes states of an electron, where $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ describes spin "up" and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ describes spin "down". Other states, like $\begin{pmatrix} i \\ -i \end{pmatrix}$ are permissible, since the base field is the complex numbers.

Complex numbers are extremely useful because of a special property that they enjoy: every polynomial over the complex numbers factors into a product of linear polynomials. For example, the polynomial $x^2 + 1$ doesn't factor over the real numbers, but over the complex numbers it factors into $(x+i)(x-i)$. This property ends up having very far-reaching consequences: often in mathematics problems that are very difficult when working over the real numbers become relatively simple when working over the complex numbers. One example of this phenomenon occurs when diagonalizing matrices, which we will learn about later in the course.

Another useful field is the rational numbers $\mathbb{Q}$. This is field is important in computer algebra: a real number given by an infinite string of numbers after the decimal point can't be stored by a computer. So instead rational approximations are used. Since the rationals are a field, the mathematics of vector spaces still apply to this special case.

There are many other examples of fields, including fields with only finitely many numbers. One example of this is the field $\mathbb{Z}_2$ which only has elements $\{0, 1\}$. Multiplication is defined normally, and addition is the usual addition, but with

$$1 + 1 = 0\,.$$

This particular field has important applications in computer science: Modern computers actually use $\mathbb{Z}_2$ arithmetic for every operation.

In fact, for every prime number p, the set $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$ forms a field. The addition and multiplication are obtained by using the usual operations over the integers, and then dividing by p and taking the remainder. For example, in $\mathbb{Z}_5$, we have $4+3 = 2$, and $4 \cdot 4 = 1$. (This is sometimes called "clock arithmetic.") Such fields are very important in computer science, cryptography, and number theory.

In this class, we will work mainly over the Real numbers and the Complex numbers, and occasionally work over $\mathbb{Z}_2$. The full story of fields is typically covered in a class on abstract algebra or Galois theory.

References

Hefferon, Chapter One, Section I.1
Beezer, Chapter VS, Section VS
Wikipedia:

- Vector Space
- Field
- Spin $\frac{1}{2}$

1. Check that $V = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} : x, y \in \mathbb{R} \right\} = \mathbb{R}^2$ with the usual addition and scalar multiplication is a vector space.

2. Check that the complex numbers form a vector space.

3. (a) Consider the set of convergent sequences, with the same addition and scalar multiplication that we defined for the space of sequences:
$$V = \left\{ f | f \colon \mathbb{N} \to \mathbb{R}, \lim_{n\to\infty} f \in \mathbb{R} \right\}$$
Is this still a vector space? Explain why or why not.

   (b) Now consider the set of divergent sequences, with the same addition and scalar multiplication as before:
$$V = \left\{ f | f \colon \mathbb{N} \to \mathbb{R}, \lim_{n\to\infty} f \text{ does not exist or is } \pm\infty \right\}$$
Is this a vector space? Explain why or why not.

4. Consider the set of 2×4 matrices:
$$V = \left\{ \begin{pmatrix} a & b & c & d \\ e & f & g & h \end{pmatrix} | a, b, c, d, e, f, g, h \in \mathbb{C} \right\}$$
Propose definitions for addition and scalar multiplication in V. Identify the zero vector in V, and check that every matrix has an additive inverse.

5. Let $P_3^{\mathbb{R}}$ be the set of polynomials with real coefficients of degree three or less.

   - Propose a definition of addition and scalar multiplication to make $P_3^{\mathbb{R}}$ a vector space.
   - Identify the zero vector, and find the additive inverse for the vector $-3 - 2x + x^2$.
   - Show that $P_3^{\mathbb{R}}$ is not a vector space over $\mathbb{C}$. Propose a small change to the definition of $P_3^{\mathbb{R}}$ to make it a vector space over $\mathbb{C}$.

7 Linear Transformations

Recall that the key properties of vector spaces are vector addition and scalar multiplication. Now suppose we have two vector spaces V and W and a map L between them:

$$L\colon V \to W$$

Now, both V and W have notions of vector addition and scalar multiplication. It would be ideal if the map L *preserved* these operations. In other words, if adding vectors and then applying L were the same as applying L to two vectors and then adding them. Likewise, it would be nice if, when multiplying by a scalar, it didn't matter whether we multiplied before or after applying L. In formulas, this means that for any $u, v \in V$ and $c \in \mathbb{R}$:

$$\begin{aligned} L(u+v) &= L(u) + L(v) \\ L(cv) &= cL(v) \end{aligned}$$

Combining these two requirements into one equation, we get the definition of a linear function or linear transformation.

Definition A function $L\colon V \to W$ is linear if for all $u, v \in V$ and $r, s \in \mathbb{R}$ we have

$$L(ru + sv) = rL(u) + sL(v)$$

Notice that on the left the addition and scalar multiplication occur in V, while on the right the operations occur in W. This is often called the *linearity property* of a linear transformation.

Reading homework: problem 7.1

Example Take $L\colon \mathbb{R}^3 \to \mathbb{R}^3$ defined by:

$$L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+y \\ y+z \\ 0 \end{pmatrix}$$

Call $u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, v = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$. Now check linearity.

$$
\begin{aligned}
L(ru+sv) &= L\left(r\begin{pmatrix}x\\y\\z\end{pmatrix}+s\begin{pmatrix}a\\b\\c\end{pmatrix}\right)\\
&= L\left(\begin{pmatrix}rx\\ry\\rz\end{pmatrix}+\begin{pmatrix}sa\\sb\\sc\end{pmatrix}\right)\\
&= L\begin{pmatrix}rx+sa\\ry+sb\\rz+sx\end{pmatrix}\\
&= \begin{pmatrix}rx+sa+ry+sb\\ry+sb+rz+sx\\0\end{pmatrix}
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
rL(u)+sL(v) &= rL\begin{pmatrix}x\\y\\z\end{pmatrix}+sL\begin{pmatrix}a\\b\\c\end{pmatrix}\\
&= r\begin{pmatrix}x+y\\y+z\\0\end{pmatrix}+s\begin{pmatrix}a+b\\b+c\\0\end{pmatrix}\\
&= \begin{pmatrix}rx+ry\\ry+rz\\0\end{pmatrix}+\begin{pmatrix}sa+sb\\sb+sc\\0\end{pmatrix}\\
&= \begin{pmatrix}rx+sa+ry+sb\\ry+sb+rz+sx\\0\end{pmatrix}
\end{aligned}
$$

Then the two sides of the linearity requirement are equal, so L is a linear transformation.

Remark We can write the linear transformation L in the previous example using a matrix like so:

$$
L\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}x+y\\y+z\\0\end{pmatrix}=\begin{pmatrix}1&1&0\\0&1&1\\0&0&0\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}
$$

Reading homework: problem 7.2

We previously checked that matrix multiplication on vectors obeyed the rule $M(ru + sv) = rMu + sMv$, so matrix multiplication is linear. As such, our check on L was guaranteed to work. In fact, matrix multiplication on vectors is a linear transformation.

Example Let V be the vector space of polynomials of finite degree with standard addition and scalar multiplication.

$$V = \{a_0 + a_1x + \cdots + a_nx^n | n \in \mathbb{N}, a_i \in \mathbb{R}\}$$

Let $L \colon V \to V$ be the derivative $\frac{d}{dx}$. For p_1 and p_2 polynomials, the rules of differentiation tell us that

$$\frac{d}{dx}(rp_1 + sp_2) = r\frac{dp_1}{dx} + s\frac{dp_2}{dx}$$

Thus, the derivative is a linear function from the set of polynomials to itself.

We can represent a polynomial as a "semi-infinite vector", like so:

$$a_0 + a_1x + \cdots + a_nx^n \longleftrightarrow \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

Then we have:

$$\frac{d}{dx}(a_0 + a_1x + \cdots + a_nx^n) = a_1 + 2a_2x + \cdots + na_nx^{n-1} \longleftrightarrow \begin{pmatrix} a_1 \\ 2a_2 \\ \vdots \\ na_n \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

One could then write the derivative as an "infinite matrix":

$$\frac{d}{dx} \longleftrightarrow \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 2 & 0 & \cdots \\ 0 & 0 & 0 & 3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

Foreshadowing Dimension. You probably have some intuitive notion of what dimension means, though we haven't actually defined the idea of dimension mathematically yet. Some of the examples of vector spaces we have worked with have been finite dimensional. (For example, $\mathbb{R}^n$ will turn out to have dimension n.) The polynomial example above is an example of an infinite dimensional vector space.

Roughly speaking, dimension is the number of independent directions available. To figure out dimension, I stand at the origin, and pick a direction. If there are any vectors in my vector space that aren't in that direction, then I choose another direction that isn't in the line determined by the direction I chose. If there are any vectors in my vector space not in the plane determined by the first two directions, then I choose one of them as my next direction. In other words, I choose a collection of *independent* vectors in the vector space. The size of a minimal set of independent vectors is the dimension of the vector space.

For finite dimensional vector spaces, linear transformations can always be represented by matrices. For that reason, we will start studying matrices intensively in the next few lectures.

References

Hefferon, Chapter Three, Section II. (Note that Hefferon uses the term *homomorphism* for a linear map. 'Homomorphism' is a very general term which in mathematics means 'Structure-preserving map.' A linear map preserves the linear structure of a vector space, and is thus a type of homomorphism.)

Beezer, Chapter LT, Section LT, Subsections LT, LTC, and MLT.
Wikipedia:

- Linear Transformation
- Dimension

Review Questions

1. Show that the pair of conditions:

   (i) $L(u+v) = L(u) + L(v)$

(ii) $L(cv) = cL(v)$

is equivalent to the single condition:

(iii) $L(ru + sv) = rL(u) + sL(v)$.

Your answer should have two parts. Show that (i,ii)$\Rightarrow$(iii), and then show that (iii)$\Rightarrow$(i,ii).

2. Let P_n be the space of polynomials of degree n or less in the variable t. Suppose L is a linear transformation from $P_2 \to P_3$ such that $L(1) = 4, L(t) = t^3$, and $L(t^2) = t - 1$.
   - Find $L(1 + t + 2t^2)$.
   - Find $L(a + bt + ct^2)$.
   - Find all values a, b, c such that $L(a + bt + ct^2) = 1 + 3t + 2t^3$.

3. Show that integration is a linear transformation on the vector space of polynomials. What would a matrix for integration look like? Be sure to think about what to do with the constant of integration.

8 Matrices

Definition An $r \times k$ matrix $M = (m^i_j)$ for $i = 1, \ldots, r; j = 1, \ldots, k$ is a rectangular array of real (or complex) numbers:

$$M = \begin{pmatrix} m^1_1 & m^1_2 & \cdots & m^1_k \\ m^2_1 & m^2_2 & \cdots & m^2_k \\ \vdots & \vdots & & \vdots \\ m^r_1 & m^r_2 & \cdots & m^r_k \end{pmatrix}$$

The numbers m^i_j are called *entries*. The superscript indexes the row of the matrix and the subscript indexes the column of the matrix in which m^i_j appears[2].

It is often useful to consider matrices whose entries are more general than the real numbers, so we allow that possibility.

An $r \times 1$ matrix $v = (v^r_1) = (v^r)$ is called a *column vector*, written

$$v = \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^r \end{pmatrix} .$$

A $1 \times k$ matrix $v = (v^1_k) = (v_k)$ is called a *row vector*, written

$$v = \begin{pmatrix} v_1 & v_2 & \cdots & v_k \end{pmatrix} .$$

Matrices are a very useful and efficient way to store information:

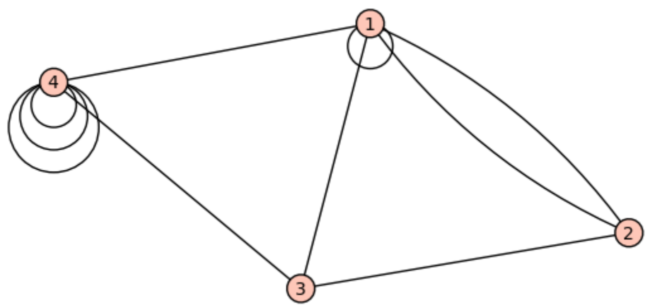
Example In computer graphics, you may have encountered image files with a .gif extension. These files are actually just matrices: at the start of the file the size of the matrix is given, and then each entry of the matrix is a number indicating the color of a particular pixel in the image.

The resulting matrix then has its rows shuffled a bit: by listing, say, every eighth row, then a web browser downloading the file can start displaying an incomplete version of the picture before the download is complete.

Finally, a compression algorithm is applied to the matrix to reduce the size of the file.

[2]This notation was first introduced by Albert Einstein.

Example Graphs occur in many applications, ranging from telephone networks to airline routes. In the subject of *graph theory*, a graph is just a collection of vertices and some edges connecting vertices. A matrix can be used to indicate how many edges attach one vertex to another.



For example, the graph pictured above would have the following matrix, where m^i_j indicates the number of edges between the vertices labeled i and j:

$$M = \begin{pmatrix} 1 & 2 & 1 & 1 \\ 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 3 \end{pmatrix}$$

This is an example of a *symmetric matrix*, since $m^i_j = m^j_i$.

The space of $r \times k$ matrices M^r_k is a vector space with the addition and scalar multiplication defined as follows:

$$\begin{aligned} M + N &= (m^i_j) + (n^i_j) = (m^i_j + n^i_j) \\ rM &= r(m^i_j) = (rm^i_j) \end{aligned}$$

In other words, addition just adds corresponding entries in two matrices, and scalar multiplication multiplies every entry. Notice that $M^n_1 = \mathbb{R}^n$ is just the vector space of column vectors.

Recall that we can multiply an $r \times k$ matrix by a $k \times 1$ column vector to produce a $r \times 1$ column vector using the rule

$$MV = \sum_{j=1}^{k} m^i_j v^j.$$

This suggests a rule for multiplying an $r \times k$ matrix M by a $k \times s$ matrix N: our $k \times s$ matrix N consists of s column vectors side-by-side, each of dimension $k \times 1$. We can multiply our $r \times k$ matrix M by each of these s column vectors using the rule we already know, obtaining s column vectors each of dimension $r \times 1$. If we place these s column vectors side-by-side, we obtain an $r \times s$ matrix MN.

That is, let

$$N = \begin{pmatrix} n_1^1 & n_2^1 & \cdots & n_s^1 \\ n_1^2 & n_2^2 & \cdots & n_s^2 \\ \vdots & \vdots & & \vdots \\ n_1^k & n_2^k & \cdots & n_s^k \end{pmatrix}$$

and call the columns N_1 through N_s:

$$N_1 = \begin{pmatrix} n_1^1 \\ n_1^2 \\ \vdots \\ n_1^k \end{pmatrix}, N_2 = \begin{pmatrix} n_2^1 \\ n_2^2 \\ \vdots \\ n_2^k \end{pmatrix}, \ldots, N_s = \begin{pmatrix} n_s^1 \\ n_s^2 \\ \vdots \\ n_s^k \end{pmatrix}.$$

Then

$$MN = M \begin{pmatrix} | & | & & | \\ N_1 & N_2 & \cdots & N_s \\ | & | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ MN_1 & MN_2 & \cdots & MN_s \\ | & | & & | \end{pmatrix}$$

A more concise way to write this rule is: If $M = (m_j^i)$ for $i = 1, \ldots, r; j = 1, \ldots, k$ and $N = (n_j^i)$ for $i = 1, \ldots, k; j = 1, \ldots, s$, then $MN = L$ where $L = (\ell_j^i)$ for $i = i, \ldots, r; j = 1, \ldots, s$ is given by

$$\ell_j^i = \sum_{p=1}^{k} m_p^i n_j^p.$$

This rule obeys linearity.

Notice that in order for the multiplication to make sense, the columns and rows must match. For an $r \times k$ matrix M and an $s \times m$ matrix N, then to make the product MN we must have $k = s$. Likewise, for the product NM, it is required that $m = r$. A common shorthand for keeping track of the sizes of the matrices involved in a given product is:

$$\left(r \times k\right) \times \left(k \times m\right) = \left(r \times m\right)$$

Example Multiplying a (3×1) matrix and a (1×2) matrix yields a (3×2) matrix.

$$\begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} \begin{pmatrix} 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 \cdot 2 & 1 \cdot 3 \\ 3 \cdot 2 & 3 \cdot 3 \\ 2 \cdot 2 & 2 \cdot 3 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 6 & 9 \\ 4 & 6 \end{pmatrix}$$

Reading homework: problem 8.1

Recall that $r \times k$ matrices can be used to represent linear transformations $\mathbb{R}^k \to \mathbb{R}^r$ via

$$MV = \sum_{j=1}^{k} m_j^i v^j,$$

which is the same rule we use when we multiply an $r \times k$ matrix by a $k \times 1$ vector to produce an $r \times 1$ vector.

Likewise, we can use a matrix $N = (n_j^i)$ to represent a linear transformation

$$L \colon M_k^s \xrightarrow{N} M_k^r$$

via

$$L(M)_l^i = \sum_{j=1}^{s} n_j^i m_l^j.$$

This is the same as the rule we use to multiply matrices. In other words, $L(M) = NM$ is a linear transformation.

Matrix Terminology The entries m_i^i are called *diagonal*, and the set $\{m_1^1, m_2^2, \ldots\}$ is called the *diagonal of the matrix*.

Any $r \times r$ matrix is called a *square matrix*. A square matrix that is zero for all non-diagonal entries is called a diagonal matrix.

The $r \times r$ diagonal matrix with all diagonal entries equal to 1 is called the *identity matrix*, I_r, or just $\mathbb{1}$. An identity matrix looks like

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

The identity matrix is special because

$$I_r M = M I_k = M$$

for all M of size $r \times k$.

In the matrix given by the product of matrices above, the diagonal entries are 2 and 9. An example of a diagonal matrix is

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Definition The *transpose* of an $r \times k$ matrix $M = (m^i_j)$ is the $k \times r$ matrix with entries

$$M^T = (\bar{m}^i_j)$$

with $\bar{m}^i_j = m^j_i$.

A matrix M is *symmetric* if $M = M^T$.

Example $\begin{pmatrix} 2 & 5 & 6 \\ 1 & 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 2 & 1 \\ 5 & 3 \\ 6 & 4 \end{pmatrix}$

Reading homework: problem 8.2

Observations

- Only square matrices can be symmetric.
- The transpose of a column vector is a row vector, and vice-versa.
- Taking the transpose of a matrix twice does nothing. *i.e.*, $(M^T)^T = M$.

Theorem 8.1 (Transpose and Multiplication)**.** *Let M, N be matrices such that MN makes sense. Then $(MN)^T = N^T M^T$.*

The proof of this theorem is left to Review Question 2.

Many properties of matrices following from the same property for real numbers. Here is an example.

Example *Associativity of matrix multiplication.* We know for real numbers x, y and z that

$$x(yz) = (xy)z\,,$$

i.e. the order of bracketing does not matter. The same property holds for matrix multiplication, let us show why. Suppose $M = \big(m^i_j\big)$, $N = \big(n^j_k\big)$ and $R = \big(r^k_l\big)$ are, respectively, $m \times n$, $n \times r$ and $r \times t$ matrices. Then from the rule for matrix multiplication we have

$$MN = \Big(\sum_{j=1}^{n} m^i_j n^j_k\Big) \text{ and } NR = \Big(\sum_{k=1}^{r} n^j_k r^k_l\Big)\,.$$

So first we compute

$$(MN)R = \Big(\sum_{k=1}^{r}\Big[\sum_{j=1}^{n} m^i_j n^j_k\Big] r^k_l\Big) = \Big(\sum_{k=1}^{r}\sum_{j=1}^{n}\Big[m^i_j n^j_k\Big] r^k_l\Big) = \Big(\sum_{k=1}^{r}\sum_{j=1}^{n} m^i_j n^j_k r^k_l\Big)\,.$$

In the first step we just wrote out the definition for matrix multiplication, in the second step we moved summation symbol outside the bracket (this is just the distributive property $x(y+z) = xy+xz$ for numbers) and in the last step we used the associativity property for real numbers to remove the square brackets. Exactly the same reasoning shows that

$$M(NR) = \Big(\sum_{j=1}^{n} m^i_j\Big[\sum_{k=1}^{r} n^j_k r^k_l\Big]\Big) = \Big(\sum_{k=1}^{r}\sum_{j=1}^{n} m^i_j\Big[n^j_k r^k_l\Big]\Big) = \Big(\sum_{k=1}^{r}\sum_{j=1}^{n} m^i_j n^j_k r^k_l\Big)\,.$$

This is the same as above so we are done. *As a fun remark, note that Einstein would simply have written* $(MN)R = (m^i_j n^j_k) r^k_l = m^i_j n^j_k r^k_l = m^i_j (n^j_k r^k_l) = M(NR)$.

References

Hefferon, Chapter Three, Section IV, parts 1-3.
Beezer, Chapter M, Section MM.
Wikipedia:

- Matrix Multiplication

Review Questions

1. Compute the following matrix products

$$\begin{pmatrix}1&2&1\\4&5&2\\7&8&2\end{pmatrix}\begin{pmatrix}-2&\frac{4}{3}&-\frac{1}{3}\\2&-\frac{5}{3}&\frac{2}{3}\\-1&2&-1\end{pmatrix}, \qquad \begin{pmatrix}1&2&3&4&5\end{pmatrix}\begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix},$$

$$\begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix}\begin{pmatrix}1&2&3&4&5\end{pmatrix}, \qquad \begin{pmatrix}1&2&1\\4&5&2\\7&8&2\end{pmatrix}\begin{pmatrix}-2&\frac{4}{3}&-\frac{1}{3}\\2&-\frac{5}{3}&\frac{2}{3}\\-1&2&-1\end{pmatrix}\begin{pmatrix}1&2&1\\4&5&2\\7&8&2\end{pmatrix},$$

$$\begin{pmatrix}x&y&z\end{pmatrix}\begin{pmatrix}2&1&1\\1&2&1\\1&1&2\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}, \qquad \begin{pmatrix}2&1&2&1&2\\0&2&1&2&1\\0&1&2&1&2\\0&2&1&2&1\\0&0&0&0&2\end{pmatrix}\begin{pmatrix}1&2&1&2&1\\0&1&2&1&2\\0&2&1&2&1\\0&1&2&1&2\\0&0&0&0&1\end{pmatrix},$$

$$\begin{pmatrix}-2&\frac{4}{3}&-\frac{1}{3}\\2&-\frac{5}{3}&\frac{2}{3}\\-1&2&-1\end{pmatrix}\begin{pmatrix}4&\frac{2}{3}&-\frac{2}{3}\\6&\frac{5}{3}&-\frac{2}{3}\\12&-\frac{16}{3}&\frac{10}{3}\end{pmatrix}\begin{pmatrix}1&2&1\\4&5&2\\7&8&2\end{pmatrix}.$$

2. Let's prove the theorem $(MN)^T = N^T M^T$.

   Note: the following is a common technique for proving matrix identities.

   (a) Let $M = (m^i_j)$ and let $N = (n^i_j)$. Write out a few of the entries of each matrix in the form given at the beginning of this chapter.

   (b) Multiply out MN and write out a few of its entries in the same form as in part a. In terms of the entries of M and the entries of N, what is the entry in row i and column j of MN?

   (c) Take the transpose $(MN)^T$ and write out a few of its entries in the same form as in part a. In terms of the entries of M and the entries of N, what is the entry in row i and column j of $(MN)^T$?

(d) Take the transposes N^T and M^T and write out a few of their entries in the same form as in part a.

(e) Multiply out $N^T M^T$ and write out a few of its entries in the same form as in part a. In terms of the entries of M and the entries of N, what is the entry in row i and column j of $N^T M^T$?

(f) Show that the answers you got in parts c and e are the same.

3. Let M be any $m \times n$ matrix. Show that $M^T M$ and MM^T are symmetric. (Hint: use the result of the previous problem.) What are their sizes?

4. Let $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ be column vectors. Show that the dot product $x \cdot y = x^T \, \mathbb{1} \, y$.

5. Above, we showed that *left* multiplication by an $r \times s$ matrix N was a linear transformation $M^s_k \xrightarrow{N} M^r_k$. Show that *right* multiplication by a $k \times m$ matrix R is a linear transformation $M^s_k \xrightarrow{R} M^s_m$. In other words, show that right matrix multiplication obeys linearity.

6. Explain what happens to a matrix when:

(a) You multiply it on the left by a diagonal matrix.

(b) You multiply it on the right by a diagonal matrix.

Give a few simple examples before you start explaining.

9 Properties of Matrices

9.1 Block Matrices

It is often convenient to partition a matrix M into smaller matrices called *blocks*, like so:

$$M = \left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 1 \\ \hline 0 & 1 & 2 & 0 \end{array}\right) = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$$

Here $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$, $B = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $C = \begin{pmatrix} 0 & 1 & 2 \end{pmatrix}$, $D = (0)$.

- The blocks of a block matrix must fit together to form a rectangle. So $\left(\begin{array}{c|c} B & A \\ \hline D & C \end{array}\right)$ makes sense, but $\left(\begin{array}{c|c} C & B \\ \hline D & A \end{array}\right)$ does not.

  Reading homework: problem 9.1

- There are many ways to cut up an $n \times n$ matrix into blocks. Often context or the entries of the matrix will suggest a useful way to divide the matrix into blocks. For example, if there are large blocks of zeros in a matrix, or blocks that look like an identity matrix, it can be useful to partition the matrix accordingly.

- Matrix operations on block matrices can be carried out by treating the blocks as matrix entries. In the example above,

$$\begin{aligned} M^2 &= \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) \\ &= \left(\begin{array}{c|c} A^2 + BC & AB + BD \\ \hline CA + DC & CB + D^2 \end{array}\right) \end{aligned}$$

Computing the individual blocks, we get:

$$
\begin{aligned}
A^2 + BC &= \begin{pmatrix} 30 & 37 & 44 \\ 66 & 81 & 96 \\ 102 & 127 & 152 \end{pmatrix} \\
AB + BD &= \begin{pmatrix} 4 \\ 10 \\ 16 \end{pmatrix} \\
CA + DC &= \begin{pmatrix} 18 \\ 21 \\ 24 \end{pmatrix} \\
CB + D^2 &= (2)
\end{aligned}
$$

Assembling these pieces into a block matrix gives:

$$
\left(\begin{array}{ccc|c} 30 & 37 & 44 & 4 \\ 66 & 81 & 96 & 10 \\ 102 & 127 & 152 & 16 \\ \hline 4 & 10 & 16 & 2 \end{array}\right)
$$

This is exactly M^2.

9.2 The Algebra of Square Matrices

Not every pair of matrices can be multiplied. When multiplying two matrices, the number of rows in the left matrix must equal the number of columns in the right. For an $r \times k$ matrix M and an $s \times l$ matrix N, then we must have $k = s$.

This is not a problem for square matrices of the same size, though. Two $n \times n$ matrices can be multiplied in either order. For a single matrix $M \in M_n^n$, we can form $M^2 = MM$, $M^3 = MMM$, and so on, and define $M^0 = I_n$, the identity matrix.

As a result, any polynomial equation can be evaluated on a matrix.

Example Let $f(x) = x - 2x^2 + 3x^3$.

Let $M = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$. Then:

$$
M^2 = \begin{pmatrix} 1 & 2t \\ 0 & 1 \end{pmatrix}, M^3 = \begin{pmatrix} 1 & 3t \\ 0 & 1 \end{pmatrix}, \ldots
$$

Hence:

$$
\begin{aligned}
f(M) &= \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} - 2\begin{pmatrix} 1 & 2t \\ 0 & 1 \end{pmatrix} + 3\begin{pmatrix} 1 & 3t \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 2 & 6t \\ 0 & 2 \end{pmatrix}
\end{aligned}
$$

Suppose $f(x)$ is any function defined by a convergent Taylor Series:

$$f(x) = f(0) + f'(0)x + \frac{1}{2!}f''(0)x^2 + \cdots$$

Then we can define the matrix function by just plugging in M:

$$f(M) = f(0) + f'(0)M + \frac{1}{2!}f''(0)M^2 + \cdots$$

There are additional techniques to determine the convergence of Taylor Series of matrices, based on the fact that the convergence problem is simple for diagonal matrices. It also turns out that $\exp(M) = 1 + M + \frac{1}{2}M^2 + \frac{1}{3!}M^3 + \cdots$ always converges.

Matrix multiplication does *not* commute. For *generic* $n \times n$ square matrices M and N, then $MN \neq NM$. For example:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

On the other hand:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

Since $n \times n$ matrices are linear transformations $\mathbb{R}^n \to \mathbb{R}^n$, we can see that the order of successive linear transformations matters. For two linear transformations K and L taking $\mathbb{R}^n \to \mathbb{R}^n$, and $v \in \mathbb{R}^n$, then in general

$$K(L(v)) \neq L(K(v))\,.$$

Finding matrices such that $MN = NM$ is an important problem in mathematics.

Trace

Matrices contain a great deal of information, so finding ways to extract essential information is useful.

Definition The *trace* of a square matrice $M = (m^i_j)$ is the sum of its diagonal entries.

$$\operatorname{tr} M = \sum_{i=1}^{n} m^i_i \, .$$

Example

$$\operatorname{tr} \begin{pmatrix} 2 & 7 & 6 \\ 9 & 5 & 1 \\ 4 & 3 & 8 \end{pmatrix} = 2 + 5 + 8 = 15$$

While matrix multiplication does not commute, the trace of a product of matrices does not depend on the order of multiplication:

$$\begin{aligned} \operatorname{tr}(MN) &= \operatorname{tr}(\sum_l M^i_l N^l_j) \\ &= \sum_i \sum_l M^i_l N^l_i \\ &= \sum_l \sum_i N^l_i M^i_l \\ &= \operatorname{tr}(\sum_i N^l_i M^i_l) \\ &= \operatorname{tr}(NM). \end{aligned}$$

Thus,

$$\operatorname{tr}(MN) = \operatorname{tr}(NM)$$

for any square matrices M and N.

In the previous example,

$$M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, N = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

$$MN = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \neq NM = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

However, $\operatorname{tr}(MN) = 2 + 1 = 3 = 1 + 2 = \operatorname{tr}(NM)$.

Another useful property of the trace is that:

$$\operatorname{tr} M = \operatorname{tr} M^T$$

This is true because the trace only uses the diagonal entries, which are fixed by the transpose. For example: $\operatorname{tr}\begin{pmatrix}1 & 1\\ 2 & 3\end{pmatrix} = 4 = \operatorname{tr}\begin{pmatrix}1 & 2\\ 1 & 3\end{pmatrix} = \operatorname{tr}\begin{pmatrix}1 & 2\\ 1 & 3\end{pmatrix}^T$

Finally, trace is a linear transformation from matrices to the real numbers. This is easy to check.

Linear Systems Redux Recall that we can view a linear system as a matrix equation

$$MX = V,$$

with M an $r \times k$ matrix of coefficients, X a $k \times 1$ matrix of unknowns, and V an $r \times 1$ matrix of constants. If M is a square matrix, then the number of equations r is the same as the number of unknowns k, so we have hope of finding a single solution.

Above we discussed functions of matrices. An extremely useful function would be $f(M) = \frac{1}{M}$, where $M\frac{1}{M} = I$. If we could compute $\frac{1}{M}$, then we would multiply both sides of the equation $MX = V$ by $\frac{1}{M}$ to obtain the solution immediately: $X = \frac{1}{M}V$.

Clearly, if the linear system has no solution, then there can be no hope of finding $\frac{1}{M}$, since if it existed we could find a solution. On the other hand, if the system has more than one solution, it also seems unlikely that $\frac{1}{M}$ would exist, since $X = \frac{1}{M}V$ yields only a single solution.

Therefore $\frac{1}{M}$ only sometimes exists. It is called the *inverse* of M, and is usually written M^{-1}.

References

Beezer: Part T, Section T
Wikipedia:

- Trace (Linear Algebra)
- Block Matrix

Review Questions

1. Let $A = \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 4 \end{pmatrix}$. Find AA^T and A^TA. What can you say about matrices MM^T and M^TM in general? Explain.

2. Compute $\exp(A)$ for the following matrices:

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

3. Suppose $ad - bc \neq 0$, and let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

   (a) Find a matrix M^{-1} such that $MM^{-1} = I$.

   (b) Explain why your result explains what you found in a previous homework exercise.

   (c) Compute $M^{-1}M$.

4. Let $M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$. Divide M into named blocks, and then multiply blocks to compute M^2.

10 Inverse Matrix

Definition A square matrix M is *invertible* (or *nonsingular*) if there exists a matrix M^{-1} such that

$$M^{-1}M = I = M^{-1}M.$$

Inverse of a 2×2 Matrix Let M and N be the matrices:

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad N = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Multiplying these matrices gives:

$$MN = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} = (ad - bc)I$$

Then $M^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$, so long as $ad - bc \neq 0$.

10.1 Three Properties of the Inverse

1. If A is a square matrix and B is the inverse of A, then A is the inverse of B, since $AB = I = BA$. Then we have the identity:

   $$(A^{-1})^{-1} = A$$

2. Notice that $B^{-1}A^{-1}AB = B^{-1}IB = I = ABB^{-1}A^{-1}$. Then:

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

   Then much like the transpose, taking the inverse of a product *reverses* the order of the product.

3. Finally, recall that $(AB)^T = B^T A^T$. Since $I^T = I$, then $(A^{-1}A)^T = A^T(A^{-1})^T = I$. Similarly, $(AA^{-1})^T = (A^{-1})^T A^T = I$. Then:

   $$(A^{-1})^T = (A^T)^{-1}$$

   As such, we could even write A^{-T} for the inverse of the transpose of A (or equivalently the transpose of the inverse).

10.2 Finding Inverses

Suppose M is a square matrix and $MX = V$ is a linear system with unique solution X_0. Since there is a unique solution, $M^{-1}V$, then the reduced row echelon form of the linear system has an identity matrix on the left:

$$\begin{pmatrix} M \mid V \end{pmatrix} \sim \begin{pmatrix} I \mid M^{-1}V \end{pmatrix}$$

Solving the linear system $MX = V$ then tells us what $M^{-1}V$ is.

To solve many linear systems at once, we can consider augmented matrices with a matrix on the right side instead of a column vector, and then apply Gaussian row reduction to the left side of the matrix. Once the identity matrix is on the left side of the augmented matrix, then the solution of each of the individual linear systems is on the right.

To compute M^{-1}, we would like M^{-1}, rather than $M^{-1}V$ to appear on the right side of our augmented matrix. This is achieved by solving the collection of systems $MX = e_k$, where e_k is the column vector of zeroes with a 1 in the kth entry. *I.e.,* the $n \times n$ identity matrix can be viewed as a bunch of column vectors $I_n = (e_1\ e_2\ \cdots e_n)$. So, putting the e_k's together into an identity matrix, we get:

$$\begin{pmatrix} M \mid I \end{pmatrix} \sim \begin{pmatrix} I \mid M^{-1}I \end{pmatrix} = \begin{pmatrix} I \mid M^{-1} \end{pmatrix}$$

Example Find $\begin{pmatrix} -1 & 2 & -3 \\ 2 & 1 & 0 \\ 4 & -2 & 5 \end{pmatrix}^{-1}$. Start by writing the augmented matrix, then apply row reduction to the left side.

$$\begin{aligned}
\left(\begin{array}{rrr|rrr} -1 & 2 & -3 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ 4 & -2 & 5 & 0 & 0 & 1 \end{array}\right) &\sim \left(\begin{array}{rrr|rrr} 1 & -2 & 3 & 1 & 0 & 0 \\ 0 & 5 & -6 & 2 & 1 & 0 \\ 0 & 6 & -7 & 4 & 0 & 1 \end{array}\right) \\
&\sim \left(\begin{array}{rrr|rrr} 1 & 0 & \frac{3}{5} & \frac{-1}{4} & \frac{2}{5} & 0 \\ 0 & 1 & \frac{-6}{5} & \frac{2}{5} & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} & \frac{-6}{5} & 1 \end{array}\right) \\
&\sim \left(\begin{array}{rrr|rrr} 1 & 0 & 0 & -5 & 4 & -3 \\ 0 & 1 & 0 & 10 & -7 & 6 \\ 0 & 0 & 1 & 8 & -6 & 5 \end{array}\right)
\end{aligned}$$

At this point, we know M^{-1} assuming we didn't goof up. However, row reduction is a lengthy and arithmetically involved process, so we should *check our answer,* by confirming that $MM^{-1} = I$ (or if you prefer $M^{-1}M = I$):

$$MM^{-1} = \begin{pmatrix} -1 & 2 & -3 \\ 2 & 1 & 0 \\ 4 & -2 & 5 \end{pmatrix} \begin{pmatrix} -5 & 4 & -3 \\ 10 & -7 & 6 \\ 8 & -6 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The product of the two matrices is indeed the identity matrix, so we're done.

Reading homework: problem 10.1

10.3 Linear Systems and Inverses

If M^{-1} exists and is known, then we can immediately solve linear systems associated to M.

Example Consider the linear system:

$$\begin{array}{rrrl} -x & +2y & -3z & = 1 \\ 2x & +y & & = 2 \\ 4x & -2y & +5z & = 0 \end{array}$$

The associated matrix equation is $MX = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$, where M is the same as in the previous section. Then:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 & 2 & -3 \\ 2 & 1 & 0 \\ 4 & -2 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -5 & 4 & -3 \\ 10 & -7 & 6 \\ 8 & -6 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ -4 \\ -4 \end{pmatrix}$$

Then $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ -4 \\ -4 \end{pmatrix}$. In summary, when M^{-1} exists, then

$$MX = V \Rightarrow X = M^{-1}V\,.$$

Reading homework: problem 10.2

10.4 Homogeneous Systems

Theorem 10.1. *A square matrix M is invertible if and only if the homogeneous system*

$$MX = 0$$

has no non-zero solutions.

Proof. First, suppose that M^{-1} exists. Then $MX = 0 \Rightarrow X = M^{-1}0 = 0$. Thus, if M is invertible, then $MX = 0$ has no non-zero solutions.

On the other hand, $MX = 0$ always has the solution $X = 0$. If no other solutions exist, then M can be put into reduced row echelon form with every variable a pivot. In this case, M^{-1} can be computed using the process in the previous section. $\square$

10.5 Bit Matrices

In computer science, information is recorded using binary strings of data. For example, the following string contains an English word:

$$011011000110100101101110011001010110000101110010$$

A *bit* is the basic unit of information, keeping track of a single one or zero. Computers can add and multiply individual bits very quickly.

Consider the set $\mathbb{Z}_2 = \{0, 1\}$ with addition and multiplication given by the following tables:

$$\begin{array}{c|cc} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \qquad \begin{array}{c|cc} \cdot & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

Notice that $-1 = 1$, since $1 + 1 = 0$.

It turns out that $\mathbb{Z}_2$ is just as good as the real or complex numbers (they are all *fields*), so we can apply all of the linear algebra we have learned thus far to matrices with $\mathbb{Z}_2$ entries. A matrix with entries in $\mathbb{Z}_2$ is sometimes called a *bit matrix*.

Example $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ is an invertible matrix over $\mathbb{Z}_2$:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

This can be easily verified by multiplying:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Application: Cryptography A very simple way to hide information is to use a substitution cipher, in which the alphabet is permuted and each letter in a message is systematically exchanged for another. For example, the ROT-13 cypher just exchanges a letter with the letter thirteen places before or after it in the alphabet. For example, HELLO becomes URYYB. Applying the algorithm again decodes the message, turning URYYB back into HELLO. Substitution ciphers are easy to break, but the basic idea can be extended to create cryptographic systems that are practically uncrackable. For example, a *one-time pad* is a system that uses a different substitution for each letter in the message. So long as a particular set of substitutions is not used on more than one message, the one-time pad is unbreakable.

English characters are often stored in computers in the ASCII format. In ASCII, a single character is represented by a string of eight bits, which we can consider as a vector in $\mathbb{Z}_2^8$ (which is like vectors in $\mathbb{R}^8$, where the entries are zeros and ones). One way to create a substitution cipher, then, is to choose an 8×8 invertible bit matrix M, and multiply each letter of the message by M. Then to decode the message, each string of eight characters would be multiplied by M^{-1}.

To make the message a bit tougher to decode, one could consider pairs (or longer sequences) of letters as a single vector in $\mathbb{Z}_2^{16}$ (or a higher-dimensional space), and then use an appropriately-sized invertible matrix.

For more on cryptography, see "The Code Book," by Simon Singh (1999, Doubleday).

You should now be ready to attempt the first sample midterm.

References

Hefferon: Chapter Three, Section IV.2
Beezer: Chapter M, Section MISLE
Wikipedia: Invertible Matrix

Review Questions

1. Find formulas for the inverses of the following matrices, when they are not singular:

   i. $\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}$

   ii. $\begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix}$

   When are these matrices singular?

2. Write down all 2×2 bit matrices and decide which of them are singular. For those which are not singular, pair them with their inverse.

3. Let M be a square matrix. Explain why the following statements are equivalent:

   i. $MX = V$ has a *unique* solution for every column vector V.

   ii. M is non-singular.

   (In general for problems like this, think about the key words:

   First, suppose that there is some column vector V such that the equation $MX = V$ has two distinct solutions. Show that M must be singular; that is, show that M can have no inverse.

   Next, suppose that there is some column vector V such that the equation $MX = V$ has no solutions. Show that M must be singular.

   Finally, suppose that M is non-singular. Show that no matter what the column vector V is, there is a unique solution to $MX = V$.)

11 LU Decomposition

Certain matrices are easier to work with than others. In this section, we will see how to write any square[3] matrix M as the product of two simpler matrices. We will write

$$M = LU\,,$$

where:

- L is *lower triangular*. This means that all entries above the main diagonal are zero. In notation, $L = (l^i_j)$ with $l^i_j = 0$ for all $j > i$.

$$L = \begin{pmatrix} l^1_1 & 0 & 0 & \cdots \\ l^2_1 & l^2_2 & 0 & \cdots \\ l^3_1 & l^3_2 & l^3_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

- U is *upper triangular*. This means that all entries below the main diagonal are zero. In notation, $U = (u^i_j)$ with $u^i_j = 0$ for all $j < i$.

$$U = \begin{pmatrix} u^1_1 & u^1_2 & u^1_3 & \cdots \\ 0 & u^2_2 & u^2_3 & \cdots \\ 0 & 0 & u^3_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$M = LU$ is called an *LU decomposition* of M.

This is a useful trick for many computational reasons. It is much easier to compute the inverse of an upper or lower triangular matrix. Since inverses are useful for solving linear systems, this makes solving any linear system associated to the matrix much faster as well. We haven't talked about determinants yet, but suffice it to say that they are important and very easy to compute for triangular matrices.

Example Linear systems associated to upper triangular matrices are very easy to solve by back substitution.

$$\left(\begin{array}{cc|c} a & b & 1 \\ 0 & c & e \end{array}\right) \Rightarrow y = \frac{e}{c}\,, \quad x = \frac{1}{a}\left(1 - \frac{be}{c}\right)$$

[3] The case where M is not square is dealt with at the end of the lecture.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & d \\ a & 1 & 0 & e \\ b & c & 1 & f \end{array}\right) \Rightarrow x = d\,, \qquad y = e - ad\,, \qquad z = f - bd - c(e - ad)$$

For lower triangular matrices, *back* substitution gives a quick solution; for upper triangular matrices, *forward* substitution gives the solution.

11.1 Using LU Decomposition to Solve Linear Systems

Suppose we have $M = LU$ and want to solve the system

$$MX = LUX = V.$$

- Step 1: Set $W = \begin{pmatrix} u \\ v \\ w \end{pmatrix} = UX$.
- Step 2: Solve the system $LW = V$. This should be simple by forward substitution since L is lower triangular. Suppose the solution to $LW = V$ is W_0.
- Step 3: Now solve the system $UX = W_0$. This should be easy by backward substitution, since U is upper triangular. The solution to this system is the solution to the original system.

We can think of this as using the matrix L to perform row operations on the matrix U in order to solve the system; this idea will come up again when we study determinants.

Reading homework: problem 11.1

Example Consider the linear system:

$$\begin{aligned} 6x + 18y + 3z &= 3 \\ 2x + 12y + z &= 19 \\ 4x + 15y + 3z &= 0 \end{aligned}$$

An LU decomposition for the associated matrix M is:

$$\begin{pmatrix} 6 & 18 & 3 \\ 2 & 12 & 1 \\ 4 & 15 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 1 & 6 & 0 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 2 & 6 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

- Step 1: Set $W = \begin{pmatrix} u \\ v \\ w \end{pmatrix} = UX$.

- Step 2: Solve the system $LW = V$:

$$\begin{pmatrix} 3 & 0 & 0 \\ 1 & 6 & 0 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} 3 \\ 19 \\ 0 \end{pmatrix}$$

  By substitution, we get $u = 1$, $v = 3$, and $w = -11$. Then

$$W_0 = \begin{pmatrix} 1 \\ 3 \\ -11 \end{pmatrix}$$

- Step 3: Solve the system $UX = W_0$.

$$\begin{pmatrix} 2 & 6 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ -11 \end{pmatrix}$$

  Back substitution gives $z = -11, y = 3$, and $x = -3$.

  Then $X = \begin{pmatrix} -3 \\ 3 \\ -11 \end{pmatrix}$, and we're done.

11.2 Finding an LU Decomposition.

For any given matrix, there are actually many different LU decompositions. However, there is a unique LU decomposition in which the L matrix has ones on the diagonal; then L is called a *lower unit triangular matrix*.

To find the LU decomposition, we'll create two sequences of matrices $L_0, L_1, \ldots$ and $U_0, U_1, \ldots$ such that at each step, $L_iU_i = M$. Each of the L_i will be lower triangular, but only the last U_i will be upper triangular.

Start by setting $L_0 = I$ and $U_0 = M$, because $L_0U_0 = M$.

Next, use the first row of U_0 to zero out the first entry of every row below it. For our running example, $U_0 = M = \begin{pmatrix} 6 & 18 & 3 \\ 2 & 12 & 1 \\ 4 & 15 & 3 \end{pmatrix}$, so the second row minus $\frac{1}{3}$ of the first row will zero out the first entry in the second row.

Likewise, the third row minus $\frac{2}{3}$ of the first row will zero out the first entry in the third row.

Set L_1 to be the lower triangular matrix whose first column is filled with the constants used to zero out the first column of M. Then $L_1 = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{2}{3} & 0 & 1 \end{pmatrix}$.

Set U_1 to be the matrix obtained by zeroing out the first column of M. Then $U_1 = \begin{pmatrix} 6 & 18 & 3 \\ 0 & 6 & 0 \\ 0 & 3 & 1 \end{pmatrix}$.

Now repeat the process by zeroing the second column of U_1 below the diagonal using the second row of U_1, and putting the corresponding entries into L_1. The resulting matrices are L_2 and U_2. For our example, $L_2 = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{2}{3} & \frac{1}{2} & 1 \end{pmatrix}$, and $U_2 = \begin{pmatrix} 6 & 18 & 3 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Since U_2 is upper-triangular, we're done. Inserting the new number into L_1 to get L_2 really is safe: the numbers in the first column don't affect the second column of U_1, since the first column of U_1 is already zeroed out.

If the matrix you're working with has more than three rows, just continue this process by zeroing out the next column below the diagonal, and repeat until there's nothing left to do.

The fractions in the L matrix are admittedly ugly. For two matrices LU, we can multiply one entire column of L by a constant λ and divide the corresponding row of U by the same constant without changing the product of the two matrices. Then:

$$
\begin{aligned}
LU &= \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{2}{3} & \frac{1}{2} & 1 \end{pmatrix} I \begin{pmatrix} 6 & 18 & 3 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{2}{3} & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{6} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 6 & 18 & 3 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 3 & 0 & 0 \\ 1 & 6 & 0 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 2 & 6 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\end{aligned}
$$

The resulting matrix looks nicer, but isn't in standard form.

Reading homework: problem 11.2

For matrices that are not square, LU decomposition still makes sense. Given an $m \times n$ matrix M, for example we could write $M = LU$ with L a square lower unit triangular matrix, and U a rectangular matrix. Then L will be an $m \times m$ matrix, and U will be an $m \times n$ matrix (of the same shape as M). From here, the process is exactly the same as for a square matrix. We create a sequence of matrices L_i and U_i that is eventually the LU decomposition. Again, we start with $L_0 = I$ and $U_0 = M$.

Example Let's find the LU decomposition of $M = U_0 = \begin{pmatrix} -2 & 1 & 3 \\ -4 & 4 & 1 \end{pmatrix}$. Since M is a 2×3 matrix, our decomposition will consist of a 2×2 matrix and a 2×3 matrix. Then we start with $L_0 = I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

The next step is to zero-out the first column of M below the diagonal. There is only one row to cancel, then, and it can be removed by subtracting 2 times the first row of M to the second row of M. Then:

$$
L_1 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \qquad U_1 = \begin{pmatrix} -2 & 1 & 3 \\ 0 & 6 & -5 \end{pmatrix}
$$

Since U_1 is upper triangular, we're done. With a larger matrix, we would just continue the process.

11.3 Block LU Decomposition

Let M be a square block matrix with square blocks X, Y, Z, W such that X^{-1} exists. Then M can be decomposed with an LDU decomposition, where D is block diagonal, as follows:

$$M = \begin{pmatrix} X & Y \\ Z & W \end{pmatrix}$$

Then:

$$M = \begin{pmatrix} I & 0 \\ ZX^{-1} & I \end{pmatrix} \begin{pmatrix} X & 0 \\ 0 & W - ZX^{-1}Y \end{pmatrix} \begin{pmatrix} I & X^{-1}Y \\ 0 & I \end{pmatrix}.$$

This can be checked explicitly simply by block-multiplying these three matrices.

Example For a 2×2 matrix, we can regard each entry as a block.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

By multiplying the diagonal matrix by the upper triangular matrix, we get the standard LU decomposition of the matrix.

References

Wikipedia:

- LU Decomposition
- Block LU Decomposition

Review Questions

1. Consider the linear system:

$$\begin{array}{lllr} x^1 & & & = v^1 \\ l^2_1 x^1 & +x^2 & & = v^2 \\ \vdots & & & \vdots \\ l^n_1 x^1 & +l^n_2 x^2 & + \cdots + x^n & = v^n \end{array}$$

i. Find x^1.

ii. Find x^2.

iii. Find x^3.

k. Try to find a formula for x^k. Don't worry about simplifying your answer.

2. Let $M = \begin{pmatrix} X & Y \\ Z & W \end{pmatrix}$ be a square $n \times n$ block matrix with W invertible.

i. If W has r rows, what size are X, Y, and Z?

ii. Find a UDL decomposition for M. In other words, fill in the stars in the following equation:

$$\begin{pmatrix} X & Y \\ Z & W \end{pmatrix} = \begin{pmatrix} I & * \\ 0 & I \end{pmatrix} \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} \begin{pmatrix} I & 0 \\ * & I \end{pmatrix}$$

12 Elementary Matrices and Determinants

Given a square matrix, is there an easy way to know when it is invertible? Answering this fundamental question is our next goal.

For small cases, we already know the answer. If M is a 1×1 matrix, then $M = (m) \Rightarrow M^{-1} = (1/m)$. Then M is invertible if and only if $m \neq 0$.

For M a 2×2 matrix, we showed in Section 10 that if $M = \begin{pmatrix} m^1_1 & m^1_2 \\ m^2_1 & m^2_2 \end{pmatrix}$, then $M^{-1} = \frac{1}{m^1_1 m^2_2 - m^1_2 m^2_1} \begin{pmatrix} m^2_2 & -m^1_2 \\ -m^2_1 & m^1_1 \end{pmatrix}$. Thus M is invertible if and only if

$$m^1_1 m^2_2 - m^1_2 m^2_1 \neq 0 \,.$$

For 2×2 matrices, this quantity is called the *determinant of M*.

$$\det M = \det \begin{pmatrix} m^1_1 & m^1_2 \\ m^2_1 & m^2_2 \end{pmatrix} = m^1_1 m^2_2 - m^1_2 m^2_1$$

Example For a 3×3 matrix, $M = \begin{pmatrix} m^1_1 & m^1_2 & m^1_3 \\ m^2_1 & m^2_2 & m^2_3 \\ m^3_1 & m^3_2 & m^3_3 \end{pmatrix}$, then (by the first review question) M is non-singular if and only if:

$$\det M = m^1_1 m^2_2 m^3_3 - m^1_1 m^2_3 m^3_2 + m^1_2 m^2_3 m^3_1 - m^1_2 m^2_1 m^3_3 + m^1_3 m^2_1 m^3_2 - m^1_3 m^2_2 m^3_1 \neq 0.$$

Notice that in the subscripts, each ordering of the numbers 1, 2, and 3 occurs exactly once. Each of these is a *permutation* of the set $\{1, 2, 3\}$.

12.1 Permutations

Consider n objects labeled 1 through n and shuffle them. Each possible shuffle is called a *permutation* σ. For example, here is an example of a permutation of 5:

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 5 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ \sigma(1) & \sigma(2) & \sigma(3) & \sigma(4) & \sigma(5) \end{bmatrix}$$

We can consider σ as a function, and write $\sigma(3) = 5$, for example. Since the top line of σ is always the same, we can omit the top line and just write:

$$\sigma = \begin{bmatrix} \sigma(1) & \sigma(2) & \sigma(3) & \sigma(4) & \sigma(5) \end{bmatrix} = \begin{bmatrix} 4 & 2 & 5 & 1 & 3 \end{bmatrix}$$

The mathematics of permutations is extensive and interesting; there are a few properties of permutations that we'll need.

- There are $n!$ permutations of n distinct objects, since there are n choices for the first object, $n-1$ choices for the second once the first has been chosen, and so on.
- Every permutation can be built up by successively swapping pairs of objects. For example, to build up the permutation $\begin{bmatrix} 3 & 1 & 2 \end{bmatrix}$ from the trivial permutation $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$, you can first swap 2 and 3, and then swap 1 and 3.
- For any given permutation σ, there is some number of swaps it takes to build up the permutation. (It's simplest to use the minimum number of swaps, but you don't have to: it turns out that *any* way of building up the permutation from swaps will have have the same parity of swaps, either even or odd.) If this number happens to be even, then σ is called an *even permutation*; if this number is odd, then σ is an *odd permutation*. In fact, $n!$ is even for all $n \geq 2$, and exactly half of the permutations are even and the other half are odd. It's worth noting that the trivial permutation (which sends $i \to i$ for every i) is an even permutation, since it uses zero swaps.

Definition The *sign function* is a function $\operatorname{sgn}(\sigma)$ that sends permutations to the set $\{-1, 1\}$, defined by:

$$\operatorname{sgn}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is even;} \\ -1 & \text{if } \sigma \text{ is odd.} \end{cases}$$

Reading homework: problem 12.1

We can use permutations to give a definition of the determinant.

Definition For an $n \times n$ matrix M, the *determinant* of M (sometimes written $|M|$) is given by:

$$\det M = \sum_{\sigma} \operatorname{sgn}(\sigma)\, m^1_{\sigma(1)} m^2_{\sigma(2)} \cdots m^n_{\sigma(n)}.$$

The sum is over all permutations of n. Each summand is a product of a single entry from each row, but with the column numbers shuffled by the permutation σ.

The last statement about the summands yields a nice property of the determinant:

Theorem 12.1. *If M has a row consisting entirely of zeros, then $m^i_{\sigma(i)} = 0$ for every σ. Then* $\det M = 0$.

Example Because there are many permutations of n, writing the determinant this way for a general matrix gives a very long sum. For $n = 4$, there are $24 = 4!$ permutations, and for $n = 5$, there are already $120 = 5!$ permutations.

For a 4×4 matrix, $M = \begin{pmatrix} m^1_1 & m^1_2 & m^1_3 & m^1_4 \\ m^2_1 & m^2_2 & m^2_3 & m^2_4 \\ m^3_1 & m^3_2 & m^3_3 & m^3_4 \\ m^4_1 & m^4_2 & m^4_3 & m^4_4 \end{pmatrix}$, then $\det M$ is:

$$\begin{aligned} \det M &= m^1_1 m^2_2 m^3_3 m^4_4 - m^1_1 m^2_3 m^3_2 m^4_4 - m^1_1 m^2_2 m^3_4 m^4_3 \\ &- m^1_2 m^2_1 m^3_3 m^4_4 + m^1_1 m^2_3 m^3_4 m^4_2 + m^1_1 m^2_4 m^3_2 m^4_3 \\ &+ m^1_2 m^2_3 m^3_1 m^4_4 + m^1_2 m^2_1 m^3_4 m^4_3 \pm \text{16 more terms.} \end{aligned}$$

This is very cumbersome.

Luckily, it is very easy to compute the determinants of certain matrices. For example, if M is diagonal, then $M^i_j = 0$ whenever $i \neq j$. Then all summands of the determinant involving off-diagonal entries vanish, so:

$$\det M = \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} m^2_{\sigma(2)} \cdots m^n_{\sigma(n)} = m^1_1 m^2_2 \cdots m^n_n.$$

Thus, the determinant of a diagonal matrix is just the product of its diagonal entries.

Since the identity matrix is diagonal with all diagonal entries equal to one, we have:

$$\det I = 1.$$

We would like to use the determinant to decide whether a matrix is invertible or not. Previously, we computed the inverse of a matrix by applying row operations. As such, it makes sense to ask what happens to the determinant when row operations are applied to a matrix.

Swapping rows Swapping rows i and j (with $i < j$) in a matrix changes the determinant. For a permutation σ, let $\hat{\sigma}$ be the permutation obtained by swapping positions i and j. The sign of $\hat{\sigma}$ is the opposite of the sign of

σ. Let M be a matrix, and M' be the same matrix, but with rows i and j swapped. Then the determinant of M' is:

$$\begin{aligned}
\det M' &= \sum_\sigma \operatorname{sgn}(\sigma)\, m^1_{\sigma(1)} \cdots m^j_{\sigma(i)} \cdots m^i_{\sigma(j)} \cdots m^n_{\sigma(n)} \\
&= \sum_\sigma \operatorname{sgn}(\sigma)\, m^1_{\sigma(1)} \cdots m^i_{\sigma(j)} \cdots m^j_{\sigma(i)} \cdots m^n_{\sigma(n)} \\
&= \sum_\sigma (-\operatorname{sgn}(\hat\sigma))\, m^1_{\hat\sigma(1)} \cdots m^i_{\hat\sigma(j)} \cdots m^j_{\hat\sigma(i)} \cdots m^n_{\hat\sigma(n)} \\
&= -\sum_{\hat\sigma} \operatorname{sgn}(\hat\sigma)\, m^1_{\hat\sigma(1)} \cdots m^i_{\hat\sigma(j)} \cdots m^j_{\hat\sigma(i)} \cdots m^n_{\hat\sigma(n)} \\
&= -\det M.
\end{aligned}$$

Thus we see that swapping rows changes the sign of the determinant. *I.e.*

$$\det S^i_j M = -\det M\,.$$

Reading homework: problem 12.2

Applying this result to $M = I$ (the identity matrix) yields

$$\det S^i_j = -1\,.$$

This implies another nice property of the determinant. If two rows of the matrix are identical, then swapping the rows changes the sign of the matrix, but leaves the matrix unchanged. Then we see the following:

Theorem 12.2. *If M has two identical rows, then* $\det M = 0$.

12.2 Elementary Matrices

Our next goal is to find matrices that emulate the Gaussian row operations on a matrix. In other words, for any matrix M, and a matrix M' equal to M after a row operation, we wish to find a matrix R such that $M' = RM$.

We will first find a matrix that, when it multiplies a matrix M, rows i and j of M are swapped.

Let R^1 through R^n denote the rows of M, and let M' be the matrix M with rows i and j swapped. Then M and M' can be regarded as a block matrices:

$$M = \begin{pmatrix} \vdots \\ R^i \\ \vdots \\ R^j \\ \vdots \end{pmatrix}, \text{ and } M' = \begin{pmatrix} \vdots \\ R^j \\ \vdots \\ R^i \\ \vdots \end{pmatrix}.$$

Then notice that:

$$M' = \begin{pmatrix} \\ \vdots \\ R^j \\ \vdots \\ R^i \\ \vdots \\ \\ \end{pmatrix} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \begin{pmatrix} \\ \vdots \\ R^i \\ \vdots \\ R^j \\ \vdots \\ \\ \end{pmatrix}$$

The matrix is just the identity matrix with rows i and j swapped. This is called an *elementary matrix* E^i_j. Then, symbolically,

$$M' = E^i_j M$$

Because $\det I = 1$ and swapping a pair of rows changes the sign of the determinant, we have found that

$$\det E^i_j = -1$$

References

Hefferon, Chapter Four, Section I.1 and I.3
Beezer, Chapter D, Section DM, Subsection EM
Beezer, Chapter D, Section PDM
Wikipedia:

- Determinant
- Permutation
- Elementary Matrix

Review Questions

1. Let $M = \begin{pmatrix} m^1_1 & m^1_2 & m^1_3 \\ m^2_1 & m^2_2 & m^2_3 \\ m^3_1 & m^3_2 & m^3_3 \end{pmatrix}$. Use row operations to put M into *row echelon form*. For simplicity, assume that $m^1_1 \neq 0 \neq m^1_1 m^2_2 - m^2_1 m^1_2$.

   Prove that M is non-singular if and only if:

   $$m^1_1 m^2_2 m^3_3 - m^1_1 m^2_3 m^3_2 + m^1_2 m^2_3 m^3_1 - m^1_2 m^2_1 m^3_3 + m^1_3 m^2_1 m^3_2 - m^1_3 m^2_2 m^3_1 \neq 0$$

2. (a) What does the matrix $E^1_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ do to $M = \begin{pmatrix} a & b \\ d & c \end{pmatrix}$ under left multiplication? What about right multiplication?

   (b) Find elementary matrices $R^1(\lambda)$ and $R^2(\lambda)$ that respectively multiply rows 1 and 2 of M by λ but otherwise leave M the same under left multiplication.

   (c) Find a matrix $S^1_2(\lambda)$ that adds a multiple λ of row 2 to row 1 under left multiplication.

3. Let M be a matrix and $S^i_j M$ the same matrix with rows i and j switched. Explain every line of the series of equations proving that $\det M = -\det(S^i_j M)$.

4. The *inversion number* of a permutation σ is the number of pairs $i < j$ such that $\sigma(i) > \sigma(j)$; it's the number of "numbers that appear left of smaller numbers" in the permutation. For example, for the permutation $\sigma = [4, 2, 3, 1]$, the inversion number is 5. 4 comes before $2, 3$, and 1, and 2 and 3 both come before 1.

   One way to compute the sign of a permutation is by using the following fact:

   $$(-1)^N = \operatorname{sgn}(\sigma),$$

   where σ is a permutation with N inversions. Let's see why this is true.

   (a) What is the inversion number of the permutation $\mu = [1, 2, 4, 3]$ that exchanges 4 and 3 and leaves everything else alone? Is it an even or an odd permutation?

What is the inversion number of the permutation $\rho = [4, 2, 3, 1]$ that exchanges 1 and 4 and leaves everything else alone? Is it an even or an odd permutation?

What is the inversion number of the permutation $\tau_{i,j}$ that exchanges i and j and leaves everything else alone? Is $\tau_{i,j}$ an even or an odd permutation? If $\tau_{i,j}^2$ refers to the permutation obtained by exchanging i and j, and then exchanging i and j again, what is $\tau_{i,j}^2$?

(b) Given a permutation σ, we can make a new permutation $\tau_{i,j}\sigma$ by exchanging the ith and jth entries of σ.

What is the inversion number of the permutation $\tau_{1,3}\mu$ where μ is as in part i? Compare the parity[4] of μ to the parity of $\tau_{1,3}\mu$.

What is the inversion number of the permutation $\tau_{2,4}\rho$ where ρ is as in part i? Compare the parity of ρ to the parity of $\tau_{2,4}\rho$.

What is the inversion number of the permutation $\tau_{3,4}\rho$ where ρ is as in part i? Compare the parity of ρ to the parity of $\tau_{3,4}\rho$.

If σ has N inversions and $\tau_{i,j}\sigma$ has M inversions, *show* that N and M have different parity. In other words, applying a transposition to σ changes the number of inversions by an odd number.

(c) (Extra credit) Show that $(-1)^N = \operatorname{sgn}(\sigma)$, where σ is a permutation with N inversions. *(Hint: How many inversions does the identity permutation have? Also, recall that σ can be built up with transpositions.)*

[4]The *parity* of an integer refers to whether the integer is even or odd.

13 Elementary Matrices and Determinants II

In Lecture 12, we saw the definition of the determinant and derived an elementary matrix that exchanges two rows of a matrix. Next, we need to find elementary matrices corresponding to the other two row operations; multiplying a row by a scalar, and adding a multiple of one row to another. As a consequence, we will derive some important properties of the determinant.

Consider $M = \begin{pmatrix} R^1 \\ \vdots \\ R^n \end{pmatrix}$, where R^i are row vectors. Let $R^i(\lambda)$ be the identity matrix, with the ith diagonal entry replaced by λ, not to be confused with the row vectors. *I.e.*

$$R^i(\lambda) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}.$$

Then:

$$M' = R^i(\lambda)M = \begin{pmatrix} R^1 \\ \vdots \\ \lambda R^i \\ \vdots \\ R^n \end{pmatrix}$$

What effect does multiplication by $R^i(\lambda)$ have on the determinant?

$$\begin{aligned} \det M' &= \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} \cdots \lambda m^i_{\sigma(i)} \cdots m^n_{\sigma(n)} \\ &= \lambda \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} \cdots m^i_{\sigma(i)} \cdots m^n_{\sigma(n)} \\ &= \lambda \det M \end{aligned}$$

Thus, multiplying a row by λ multiplies the determinant by λ. *I.e.,*

$$\det R^i(\lambda)M = \lambda \det M\,.$$

Since $R^i(\lambda)$ is just the identity matrix with a single row multiplied by λ, then by the above rule, the determinant of $R^i(\lambda)$ is λ. Thus:

$$\det R^i(\lambda) = \det \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} = \lambda$$

The final row operation is adding λR^j to R^i. This is done with the matrix $S^i_j(\lambda)$, which is an identity matrix but with a λ in the i,j position.

$$S^i_j(\lambda) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & \lambda & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

Then multiplying $S^i_j(\lambda)$ by M gives the following:

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & \lambda & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \begin{pmatrix} \vdots \\ R^i \\ \vdots \\ R^j \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ R^i + \lambda R^j \\ \vdots \\ R^j \\ \vdots \end{pmatrix}$$

What is the effect of multiplying by $S^i_j(\lambda)$ on the determinant? Let $M' = S^i_j(\lambda)M$, and let M'' be the matrix M but with R^i replaced by R^j.

$$
\begin{aligned}
\det M' &= \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} \cdots (m^i_{\sigma(i)} + \lambda m^j_{\sigma(j)}) \cdots m^n_{\sigma(n)} \\
&= \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} \cdots m^i_{\sigma(i)} \cdots m^n_{\sigma(n)} \\
&\qquad + \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} \cdots \lambda m^j_{\sigma(j)} \cdots m^j_{\sigma(j)} \cdots m^n_{\sigma(n)} \\
&= \det M + \lambda \det M''
\end{aligned}
$$

Since M'' has two identical rows, its determinant is 0. Then

$$\det S^i_j(\lambda) M = \det M\,.$$

Notice that if M is the identity matrix, then we have

$$\det S^i_j(\lambda) = \det(S^i_j(\lambda) I) = \det I = 1\,.$$

We now have elementary matrices associated to each of the row operations.

$$
\begin{aligned}
E^i_j &= I \text{ with rows } i,j \text{ swapped;} \quad \det E^i_j = -1 \\
R^i(\lambda) &= I \text{ with } \lambda \text{ in position } i,i; \quad \det R^i_j(\lambda) = \lambda \\
S^i_j(\lambda) &= I \text{ with } \lambda \text{ in position } i,j; \quad \det S^i_j(\lambda) = 1
\end{aligned}
$$

We have also proved the following theorem along the way:

Theorem 13.1. *If E is* any *of the elementary matrices $E^i_j, R^i(\lambda), S^i_j(\lambda)$, then $\det(EM) = \det E \det M$.*

Reading homework: problem 13.1

We have seen that any matrix M can be put into reduced row echelon form via a sequence of row operations, and we have seen that any row operation can be emulated with left matrix multiplication by an elementary matrix. Suppose that $\mathrm{RREF}(M)$ is the reduced row echelon form of M. Then $\mathrm{RREF}(M) = E_1 E_2 \cdots E_k M$ where each E_i is an elementary matrix.

What is the determinant of a square matrix in reduced row echelon form?

- If M is not invertible, then some row of $\mathrm{RREF}(M)$ contains only zeros. Then we can multiply the zero row by any constant λ without changing M; by our previous observation, this scales the determinant of M by λ. Thus, if M is not invertible, $\det \mathrm{RREF}(M) = \lambda \det \mathrm{RREF}(M)$, and so $\det \mathrm{RREF}(M) = 0$.

- Otherwise, every row of $\mathrm{RREF}(M)$ has a pivot on the diagonal; since M is square, this means that $\mathrm{RREF}(M)$ is the identity matrix. Then if M is invertible, $\det \mathrm{RREF}(M) = 1$.

- Additionally, notice that $\det \mathrm{RREF}(M) = \det(E_1 E_2 \cdots E_k M)$. Then by the theorem above, $\det \mathrm{RREF}(M) = \det(E_1) \cdots \det(E_k) \det M$. Since each E_i has non-zero determinant, then $\det \mathrm{RREF}(M) = 0$ if and only if $\det M = 0$.

Then we have shown:

Theorem 13.2. *For any square matrix M, $\det M \neq 0$ if and only if M is invertible.*

Since we know the determinants of the elementary matrices, we can immediately obtain the following:

Corollary 13.3. *Any elementary matrix $E^i_j, R^i(\lambda), S^i_j(\lambda)$ is invertible, except for $R^i(0)$. In fact, the inverse of an elementary matrix is another elementary matrix.*

To obtain one last important result, suppose that M and N are square $n \times n$ matrices, with reduced row echelon forms such that, for elementary matrices E_i and F_i,

$$M = E_1 E_2 \cdots E_k \ \mathrm{RREF}(M)\,,$$

and

$$N = F_1 F_2 \cdots F_l \ \mathrm{RREF}(N) = N\,.$$

If $\mathrm{RREF}(M)$ is the identity matrix (ie, M is invertible), then:

$$\begin{array}{rcl}
\det(MN) & = & \det(E_1 E_2 \cdots E_k \ \mathrm{RREF}(M) F_1 F_2 \cdots F_l \ \mathrm{RREF}(N)) \\
& = & \det(E_1 E_2 \cdots E_k I F_1 F_2 \cdots F_l \ \mathrm{RREF}(N)) \\
& = & \det(E_1) \cdots \det(E_k) \det(I) \det(F_1) \cdots \det(F_l) \det(\mathrm{RREF}(N) \\
& = & \det(M) \det(N)
\end{array}$$

Otherwise, M is not invertible, and $\det M = 0 = \det \text{RREF}(M)$. Then there exists a row of zeros in $\text{RREF}(M)$, so $R^n(\lambda)\,\text{RREF}(M) = \text{RREF}(M)$. Then:

$$\begin{aligned}
\det(MN) &= \det(E_1E_2\cdots E_k\ \text{RREF}(M)N) \\
&= \det(E_1E_2\cdots E_k\ \text{RREF}(M)N) \\
&= \det(E_1)\cdots\det(E_k)\det(\text{RREF}(M)N) \\
&= \det(E_1)\cdots\det(E_k)\det(R^n(\lambda)\ \text{RREF}(M)N) \\
&= \det(E_1)\cdots\det(E_k)\lambda\det(\text{RREF}(M)N) \\
&= \lambda\det(MN)
\end{aligned}$$

Which implies that $\det(MN) = 0 = \det M \det N$.

Thus we have shown that for *any* matrices M and N,

$$\det(MN) = \det M \det N$$

This result is *extremely important*; do not forget it!

Reading homework: problem 13.2

References

Hefferon, Chapter Four, Section I.1 and I.3
Beezer, Chapter D, Section DM, Subsection EM
Beezer, Chapter D, Section PDM
Wikipedia:

- Determinant
- Elementary Matrix

Review Questions

1. Let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $N = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$. Compute the following:

   (a) $\det M$.

   (b) $\det N$.

   (c) $\det(MN)$.

(d) $\det M \det N$.

(e) $\det(M^{-1})$ assuming $ad - bc \neq 0$.

(f) $\det(M^T)$

(g) $\det(M+N) - (\det M + \det N)$. Is the determinant a linear transformation from square matrices to real numbers? Explain.

2. Suppose $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible. Write M as a product of elementary row matrices times $\text{RREF}(M)$.

3. Find the inverses of each of the elementary matrices, $E^i_j, R^i(\lambda), S^i_j(\lambda)$. Make sure to show that the elementary matrix times its inverse is actually the identity.

14 Properties of the Determinant

In Lecture 13 we showed that the determinant of a matrix is non-zero if and only if that matrix is invertible. We also showed that the determinant is a *multiplicative* function, in the sense that $\det(MN) = \det M \det N$. Now we will devise some methods for calculating the determinant.

Recall that:

$$\det M = \sum_{\sigma} \operatorname{sgn}(\sigma) m^1_{\sigma(1)} m^2_{\sigma(2)} \cdots m^n_{\sigma(n)}.$$

A *minor* of an $n \times n$ matrix M is the determinant of any square matrix obtained from M by deleting rows and columns. In particular, any entry m^i_j of a square matrix M is associated to a minor obtained by deleting the ith row and jth column of M.

It is possible to write the determinant of a matrix in terms of its minors as follows:

$$\begin{aligned}
\det M &= \sum_{\sigma} \operatorname{sgn}(\sigma)\, m^1_{\sigma(1)} m^2_{\sigma(2)} \cdots m^n_{\sigma(n)} \\
&= m^1_1 \sum_{\hat\sigma} \operatorname{sgn}(\hat\sigma)\, m^2_{\hat\sigma(2)} \cdots m^n_{\hat\sigma(n)} \\
&- m^1_2 \sum_{\hat\sigma} \operatorname{sgn}(\hat\sigma)\, m^2_{\hat\sigma(1)} m^3_{\hat\sigma(3)} \cdots m^n_{\hat\sigma(n)} \\
&+ m^1_3 \sum_{\hat\sigma} \operatorname{sgn}(\hat\sigma)\, m^2_{\hat\sigma(1)} m^3_{\hat\sigma(2)} m^4_{\hat\sigma(4)} \cdots m^n_{\hat\sigma(n)} \pm \cdots
\end{aligned}$$

Here the symbols $\hat\sigma$ refer to permutations of $n-1$ objects. What we're doing here is collecting up all of the terms of the original sum that contain the first row entry m^1_j for each column number j. Each term in that collection is associated to a permutation sending $1 \to j$. The remainder of any such permutation maps the set $\{2, \ldots, n\} \to \{1, \ldots, j-1, j+1, \ldots, n\}$. We call this partial permutation $\hat\sigma = \begin{bmatrix} \sigma(2) & \cdots & \sigma(n) \end{bmatrix}$.

The last issue is that the permutation $\hat\sigma$ may not have the same sign as σ. From previous homework, we know that a permutation has the same parity as its inversion number. Removing $1 \to j$ from a permutation reduces the inversion number by the number of elements right of j that are less than j. Since j comes first in the permutation $\begin{bmatrix} j & \sigma(2) & \cdots & \sigma(n) \end{bmatrix}$, the inversion number of $\hat{\sigma}$ is reduced by $j-1$. Then the sign of σ differs from the sign of $\hat{\sigma}$ if σ sends 1 to an even number.

In other words, to expand by minors we pick an entry m^1_j of the first row, then add $(-1)^{j-1}$ times the determinant of the matrix with row i and column j deleted.

Example Let's compute the determinant of $M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$ using expansion by minors.

$$\begin{aligned} \det M &= 1\det\begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} - 2\det\begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix} + 3\det\begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix} \\ &= 1(5\cdot 9 - 8\cdot 6) - 2(4\cdot 9 - 7\cdot 6) + 3(4\cdot 8 - 7\cdot 5) \\ &= 0 \end{aligned}$$

Here, M^{-1} does not exist because[5] $\det M = 0$.

Example Sometimes the entries of a matrix allow us to simplify the calculation of the determinant. Take $N = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 0 & 0 \\ 7 & 8 & 9 \end{pmatrix}$. Notice that the second row has many zeros; then we can switch the first and second rows of N to get:

$$\begin{aligned} \det\begin{pmatrix} 1 & 2 & 3 \\ 4 & 0 & 0 \\ 7 & 8 & 9 \end{pmatrix} &= -\det\begin{pmatrix} 4 & 0 & 0 \\ 1 & 2 & 3 \\ 7 & 8 & 9 \end{pmatrix} \\ &= -4\det\begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix} \\ &= 24 \end{aligned}$$

Theorem 14.1. *For any square matrix M, we have:*

$$\det M^T = \det M$$

[5] A fun exercise is to compute the determinant of a 4×4 matrix filled in order, from left to right, with the numbers $1, 2, 3, \dots 16$. What do you observe? Try the same for a 5×5 matrix with $1, 2, 3 \dots 25$. Is there a pattern? Can you explain it?

Proof. By definition,

$$\det M = \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} m^2_{\sigma(2)} \cdots m^n_{\sigma(n)}.$$

For any permutation σ, there is a unique inverse permutation σ^{-1} that undoes σ. If σ sends $i \to j$, then σ^{-1} sends $j \to i$. In the two-line notation for a permutation, this corresponds to just flipping the permutation over. For example, if $\sigma = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}$, then we can find σ^{-1} by flipping the permutation and then putting the columns in order:

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

Since any permutation can be built up by transpositions, one can also find the inverse of a permutation σ by undoing each of the transpositions used to build up σ; this shows that one can use the same number of transpositions to build σ and σ^{-1}. In particular, $\operatorname{sgn} \sigma = \operatorname{sgn} \sigma^{-1}$.

Reading homework: problem 14.1

Then we can write out the above in formulas as follows:

$$\begin{aligned}
\det M &= \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} m^2_{\sigma(2)} \cdots m^n_{\sigma(n)} \\
&= \sum_\sigma \operatorname{sgn}(\sigma) m_1^{\sigma^{-1}(1)} m_2^{\sigma^{-1}(2)} \cdots m_n^{\sigma^{-1}(n)} \\
&= \sum_\sigma \operatorname{sgn}(\sigma^{-1}) m_1^{\sigma^{-1}(1)} m_2^{\sigma^{-1}(2)} \cdots m_n^{\sigma^{-1}(n)} \\
&= \sum_\sigma \operatorname{sgn}(\sigma) m_1^{\sigma(1)} m_2^{\sigma(2)} \cdots m_n^{\sigma(n)} \\
&= \det M^T.
\end{aligned}$$

The second-to-last equality is due to the existence of a unique inverse permutation: summing over permutations is the same as summing over all inverses of permutations. The final equality is by the definition of the transpose. □

Example Because of this theorem, we see that expansion by minors also works over columns. Let $M = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 8 & 9 \end{pmatrix}$. Then

$$\det M = \det M^T = 1 \det \begin{pmatrix} 5 & 8 \\ 6 & 9 \end{pmatrix} = -3\,.$$

14.1 Determinant of the Inverse

Let M and N be $n \times n$ matrices. We previously showed that

$$\det(MN) = \det M \det N, \text{ and } \det I = 1.$$

Then $1 = \det I = \det(MM^{-1}) = \det M \det M^{-1}$. As such we have:

Theorem 14.2.

$$\det M^{-1} = \frac{1}{\det M}$$

14.2 Adjoint of a Matrix

Recall that for the 2×2 matrix $M = \begin{pmatrix} m^1_1 & m^1_2 \\ m^2_1 & m^2_2 \end{pmatrix}$, then

$$M^{-1} = \frac{1}{m^1_1 m^2_2 - m^1_2 m^2_1} \begin{pmatrix} m^2_2 & -m^1_2 \\ -m^2_1 & m^1_1 \end{pmatrix}\,.$$

This matrix $\begin{pmatrix} m^2_2 & -m^1_2 \\ -m^2_1 & m^1_1 \end{pmatrix}$ that appears above is a special matrix, called the *adjoint* of M. Let's define the adjoint for an $n \times n$ matrix.

A *cofactor* of M is obtained choosing any entry m^i_j of M and then deleting the ith row and jth column of M, taking the determinant of the resulting matrix, and multiplying by$(-1)^{i+j}$. This is written cofactor(m^i_j).

Definition For $M = (m^i_j)$ a square matrix, The *adjoint matrix* $\operatorname{adj} M$ is given by:

$$\operatorname{adj} M = (\text{cofactor}(m^i_j))^T$$

Example

$$\operatorname{adj}\begin{pmatrix}3&-1&-1\\1&2&0\\0&1&1\end{pmatrix}=\begin{pmatrix}\det\begin{pmatrix}2&0\\1&1\end{pmatrix}&-\det\begin{pmatrix}1&0\\0&1\end{pmatrix}&\det\begin{pmatrix}1&2\\0&1\end{pmatrix}\\-\det\begin{pmatrix}-1&-1\\1&1\end{pmatrix}&\det\begin{pmatrix}3&-1\\0&1\end{pmatrix}&-\det\begin{pmatrix}3&-1\\0&1\end{pmatrix}\\\det\begin{pmatrix}-1&-1\\2&0\end{pmatrix}&-\det\begin{pmatrix}3&-1\\1&0\end{pmatrix}&\det\begin{pmatrix}3&-1\\1&2\end{pmatrix}\end{pmatrix}^T$$

Reading homework: problem 14.2

Let's multiply $M \operatorname{adj} M$. For any matrix N, the i,j entry of MN is given by taking the dot product of the ith row of M and the jth column of N. Notice that the dot product of the ith row of M and the ith column of $\operatorname{adj} M$ is just the expansion by minors of $\det M$ in the ith row. Further, notice that the dot product of the ith row of M and the jth column of $\operatorname{adj} M$ with $j \neq i$ is the same as expanding M by minors, but with the jth row replaced by the ith row. Since the determinant of any matrix with a row repeated is zero, then these dot products are zero as well.

We know that the i,j entry of the product of two matrices is the dot product of the ith row of the first by the jth column of the second. Then:

$$M \operatorname{adj} M = (\det M)I$$

Thus, when $\det M \neq 0$, the adjoint gives an explicit formula for M^{-1}.

Theorem 14.3. *For M a square matrix with $\det M \neq 0$ (equivalently, if M is invertible), then*

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

Example Continuing with the previous example,

$$\operatorname{adj}\begin{pmatrix}3&-1&-1\\1&2&0\\0&1&1\end{pmatrix}=\begin{pmatrix}2&0&2\\-1&3&-1\\1&-3&7\end{pmatrix}.$$

Now, multiply:

$$
\begin{pmatrix} 3 & -1 & -1 \\ 1 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 2 \\ -1 & 3 & -1 \\ 1 & -3 & 7 \end{pmatrix} = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix}
$$

$$
\Rightarrow \begin{pmatrix} 3 & -1 & -1 \\ 1 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix}^{-1} = \frac{1}{6} \begin{pmatrix} 2 & 0 & 2 \\ -1 & 3 & -1 \\ 1 & -3 & 7 \end{pmatrix}
$$

This process for finding the inverse matrix is sometimes called *Cramer's Rule* .

14.3 Application: Volume of a Parallelepiped

Given three vectors u, v, w in $\mathbb{R}^3$, the parallelepiped determined by the three vectors is the "squished" box whose edges are parallel to u, v, and w as depicted in Figure 1.

From calculus, we know that the volume of this object is $|u \cdot (v \times w)|$. This is the same as expansion by minors of the matrix whose columns are u, v, w. Then:

$$
\text{Volume} = \left| \det \begin{pmatrix} u & v & w \end{pmatrix} \right|
$$

References

Hefferon, Chapter Four, Section I.1 and I.3
Beezer, Chapter D, Section DM, Subsection DD
Beezer, Chapter D, Section DM, Subsection CD
Wikipedia:

- Determinant
- Elementary Matrix
- Cramer's Rule

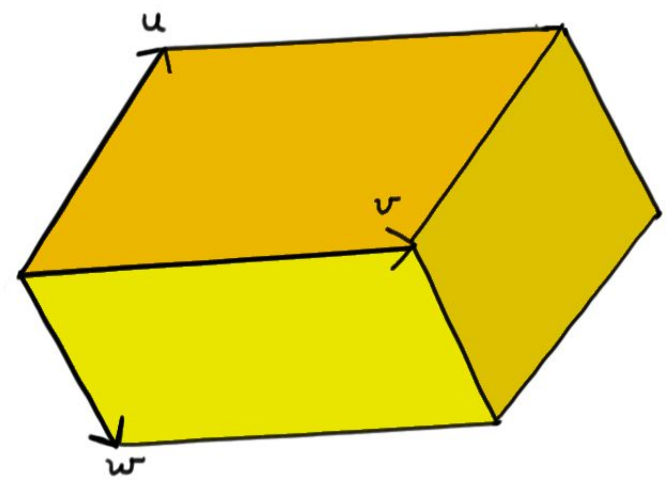


Figure 1: A parallelepiped.

Review Questions

1. Let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Show:

$$\det M = \frac{1}{2}(\operatorname{tr} M)^2 - \frac{1}{2}\operatorname{tr}(M^2)$$

   Suppose M is a 3×3 matrix. Find and verify a similar formula for $\det M$ in terms of $\operatorname{tr}(M^3)$, $(\operatorname{tr} M)(\operatorname{tr}(M^2))$, and $(\operatorname{tr} M)^3$.

2. Suppose $M = LU$ is an LU decomposition. Explain how you would efficiently compute $\det M$ in this case.

3. In computer science, the *complexity* of an algorithm is computed (roughly) by counting the number of times a given operation is performed. Suppose adding or subtracting any two numbers takes a seconds, and multiplying two numbers takes m seconds. Then, for example, computing $2 \cdot 6 - 5$ would take $a + m$ seconds.

   (a) How many additions and multiplications does it take to compute the determinant of a general 2×2 matrix?

(b) Write a formula for the number of additions and multiplications it takes to compute the determinant of a general $n \times n$ matrix using the definition of the determinant. Assume that finding and multiplying by the sign of a permutation is free.

(c) How many additions and multiplications does it take to compute the determinant of a general 3×3 matrix using expansion by minors? Assuming $m = 2a$, is this faster than computing the determinant from the definition?

15 Subspaces and Spanning Sets

It is time to study vector spaces more carefully and answer some fundamental questions.

1. *Subspaces*: When is a subset of a vector space itself a vector space? (This is the notion of a *subspace*.)

2. *Linear Independence*: Given a collection of vectors, is there a way to tell whether they are independent, or if one is a linear combination of the others?

3. *Dimension*: Is there a consistent definition of how "big" a vector space is?

4. *Basis*: How do we label vectors? Can we write any vector as a sum of some basic set of vectors? How do we change our point of view from vectors labeled one way to vectors labeled in another way?

Let's start at the top!

15.1 Subspaces

Definition We say that a subset U of a vector space V is a *subspace* of V if U is a vector space under the inherited addition and scalar multiplication operations of V.

Example Consider a plane P in $\mathbb{R}^3$ through the origin:

$$ax + by + cz = 0.$$

This equation can be expressed as the homogeneous system $\begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$, or $MX = 0$ with M the matrix $\begin{pmatrix} a & b & c \end{pmatrix}$. If X_1 and X_2 are both solutions to $MX = 0$, then, by linearity of matrix multiplication, so is $\mu X_1 + \nu X_2$:

$$M(\mu X_1 + \nu X_2) = \mu M X_1 + \nu M X_2 = 0.$$

So P is closed under addition and scalar multiplication. Additionally, P contains the origin (which can be derived from the above by setting $\mu = \nu = 0$). All other vector space requirements hold for P because they hold for all vectors in $\mathbb{R}^3$.

Theorem 15.1 (Subspace Theorem). *Let U be a non-empty subset of a vector space V. Then U is a subspace if and only if $\mu u_1 + \nu u_2 \in U$ for arbitrary u_1, u_2 in U, and arbitrary constants μ, ν.*

Proof. One direction of this proof is easy: if U is a subspace, then it is a vector space, and so by the additive closure and multiplicative closure properties of vector spaces, it has to be true that $\mu u_1 + \nu u_2 \in U$ for all u_1, u_2 in U and all constants constants μ, ν.

The other direction is almost as easy: we need to show that if $\mu u_1 + \nu u_2 \in U$ for all u_1, u_2 in U and all constants μ, ν, then U is a vector space. That is, we need to show that the ten properties of vector spaces are satisfied. We know that the additive closure and multiplicative closure properties are satisfied. Each of the other eight properties is true in U because it is true in V. The details of this are left as an exercise. □

Note that the requirements of the subspace theorem are often referred to as "closure".

From now on, we can use this theorem to check if a set is a vector space. That is, if we have some set U of vectors that come from some bigger vector space V, to check if U itself forms a smaller vector space we need check only two things: if we add any two vectors in U, do we end up with a vector in U? And, if we multiply any vector in U by any constant, do we end up with a vector in U? If the answer to both of these questions is yes, then U is a vector space. If not, U is not a vector space.

Reading homework: problem 15.1

15.2 Building Subspaces

Consider the set

$$U = \left\{ \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix} \right\} \subset \mathbb{R}^3.$$

Because U consists of only two vectors, it clear that U is *not* a vector space, since any constant multiple of these vectors should also be in U. For example, the 0-vector is not in U, nor is U closed under vector addition.

But we know that any two vectors define a plane. In this case, the vectors in U define the xy-plane in $\mathbb{R}^3$. We can consider the xy-plane as the set of all vectors that arise as a linear combination of the two vectors in U. Call this set of all linear combinations the *span* of U:

$$\operatorname{span}(U) = \left\{ x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \middle| x, y \in \mathbb{R} \right\}.$$

Notice that any vector in the xy-plane is of the form

$$\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \in \operatorname{span}(U).$$

Definition Let V be a vector space and $S = \{s_1, s_2, \ldots\} \subset V$ a subset of V. Then the *span of* S is the set:

$$\operatorname{span}(S) = \{r^1 s_1 + r^2 s_2 + \cdots + r^N s_N | r^i \in \mathbb{R}, N \in \mathbb{N}\}.$$

That is, the span of S is the set of all finite linear combinations[6] of elements of S. Any *finite* sum of the form (a constant times s_1 plus a constant times s_2 plus a constant times s_3 and so on) is in the span of S.

It is important that we only allow finite linear combinations. In the definition above, N must be a finite number. It can be any finite number, but it must be finite.

Example Let $V = \mathbb{R}^3$ and $X \subset V$ be the x-axis. Let $P = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and set

$$S = X \cup P\,.$$

The elements of $\operatorname{span}(S)$ are linear combinations of vectors in the x-axis and the vector P.

The vector $\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$ is in $\operatorname{span}(S)$, because $\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. Similarly, the vector $\begin{pmatrix} -12 \\ 17.5 \\ 0 \end{pmatrix}$ is in $\operatorname{span}(S)$, because $\begin{pmatrix} -12 \\ 17.5 \\ 0 \end{pmatrix} = \begin{pmatrix} -12 \\ 0 \\ 0 \end{pmatrix} + 17.5 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. Similarly,

[6] Usually our vector spaces are defined over $\mathbb{R}$, but in general we can have vector spaces defined over different base fields such as $\mathbb{C}$ or $\mathbb{Z}_2$. The coefficients r^i should come from whatever our base field is (usually $\mathbb{R}$).

any vector of the form

$$\begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

is in $\operatorname{span}(S)$. On the other hand, any vector in $\operatorname{span}(S)$ must have a zero in the z-coordinate. (Why?)

So $\operatorname{span}(S)$ is the xy-plane, which is a vector space. (Try drawing a picture to verify this!)

Reading homework: problem 15.2

Lemma 15.2. *For any subset $S \subset V$, $\operatorname{span}(S)$ is a subspace of V.*

Proof. We need to show that $\operatorname{span}(S)$ is a vector space.

It suffices to show that $\operatorname{span}(S)$ is closed under linear combinations. Let $u, v \in \operatorname{span}(S)$ and λ, μ be constants. By the definition of $\operatorname{span}(S)$, there are constants c^i and d^i (some of which could be zero) such that:

$$\begin{aligned} u &= c^1 s_1 + c^2 s_2 + \cdots \\ v &= d^1 s_1 + d^2 s_2 + \cdots \\ \Rightarrow \lambda u + \mu v &= \lambda(c^1 s_1 + c^2 s_2 + \cdots) + \mu(d^1 s_1 + d^2 s_2 + \cdots) \\ &= (\lambda c^1 + \mu d^1) s_1 + (\lambda c^2 + \mu d^2) s_2 + \cdots \end{aligned}$$

This last sum is a linear combination of elements of S, and is thus in $\operatorname{span}(S)$. Then $\operatorname{span}(S)$ is closed under linear combinations, and is thus a subspace of V. $\square$

Note that this proof, like many proofs, consisted of little more than just writing out the definitions.

Example For which values of a does

$$\operatorname{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ a \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}, \begin{pmatrix} a \\ 1 \\ 0 \end{pmatrix} \right\} = \mathbb{R}^3?$$

Given an arbitrary vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ in $\mathbb{R}^3$, we need to find constants r^1, r^2, r^3 such that

$$r^1 \begin{pmatrix}1\\0\\a\end{pmatrix} + r^2 \begin{pmatrix}1\\2\\-3\end{pmatrix} + r^3 \begin{pmatrix}a\\1\\0\end{pmatrix} = \begin{pmatrix}x\\y\\z\end{pmatrix}.$$

We can write this as a linear system in the unknowns r^1, r^2, r^3 as follows:

$$\begin{pmatrix}1 & 1 & a\\0 & 2 & 1\\a & -3 & 0\end{pmatrix} \begin{pmatrix}r^1\\r^2\\r^3\end{pmatrix} = \begin{pmatrix}x\\y\\z\end{pmatrix}.$$

If the matrix $M = \begin{pmatrix}1 & 1 & a\\0 & 2 & 1\\a & -3 & 0\end{pmatrix}$ is invertible, then we can find a solution

$$M^{-1} \begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}r^1\\r^2\\r^3\end{pmatrix}$$

for *any* vector $\begin{pmatrix}x\\y\\z\end{pmatrix} \in \mathbb{R}^3$.

Therefore we should choose a so that M is invertible:

$$i.e.,\ 0 \neq \det M = -2a^2 + 3 + a = -(2a-3)(a+1).$$

Then the span is $\mathbb{R}^3$ if and only if $a \neq -1, \frac{3}{2}$.

References

Hefferon, Chapter Two, Section I.2: Subspaces and Spanning Sets
Beezer, Chapter VS, Section S
Beezer, Chapter V, Section LC
Beezer, Chapter V, Section SS
Wikipedia:

- Linear Subspace
- Linear Span

Review Questions

1. (Subspace Theorem) Suppose that V is a vector space and that $U \subset V$ is a subset of V. Show that

$$\mu u_1 + \nu u_2 \in U \text{ for all } u_1, u_2 \in U, \mu, \nu \in \mathbb{R}$$

implies that U is a subspace of V. (In other words, check all the vector space requirements for U.)

2. Let $P_3^{\mathbb{R}}$ be the vector space of polynomials of degree 3 or less in the variable x. Check whether

$$x - x^3 \in \operatorname{span}\{x^2, 2x + x^2, x + x^3\}$$

3. Let U and W be subspaces of V. Are:

   (a) $U \cup W$

   (b) $U \cap W$

also subspaces? Explain why or why not. Draw examples in $\mathbb{R}^3$.

16 Linear Independence

Consider a plane P that includes the origin in $\mathbb{R}^3$ and a collection $\{u, v, w\}$ of non-zero vectors in P. If no two of u, v and w are parallel, then $P = \text{span}\{u, v, w\}$. But any two vectors determines a plane, so we should be able to span the plane using only two vectors. Then we could choose two of the vectors in $\{u, v, w\}$ whose span is P, and express the other as a linear combination of those two. Suppose u and v span P. Then there exist constants d^1, d^2 (not both zero) such that $w = d^1u + d^2v$. Since w can be expressed in terms of u and v we say that it is not independent. More generally, the relationship

$$c^1u + c^2v + c^3w = 0 \qquad c^i \in \mathbb{R}, \text{ some } c^i \neq 0$$

expresses the fact that u, v, w are not all independent.

Definition We say that the vectors $v_1, v_2, \ldots, v_n$ are *linearly dependent* if there exist constants[7] $c^1, c^2, \ldots, c^n$ not all zero such that

$$c^1v_1 + c^2v_2 + \cdots + c^nv_n = 0.$$

Otherwise, the vectors $v_1, v_2, \ldots, v_n$ are *linearly independent.*

Example Consider the following vectors in $\mathbb{R}^4$:

$$v_1 = \begin{pmatrix} 4 \\ -1 \\ 3 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} -3 \\ 7 \\ 4 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 5 \\ 12 \\ 17 \end{pmatrix}, \qquad v_4 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

Are these vectors linearly independent?

Since $3v_1 + 2v_2 - v_3 + v_4 = 0$, the vectors are linearly *dependent.*

In the above example we were given the linear combination $3v_1 + 2v_2 - v_3 + v_4$ seemingly by magic. The next example shows how to find such a linear combination, if it exists.

[7]Usually our vector spaces are defined over $\mathbb{R}$, but in general we can have vector spaces defined over different base fields such as $\mathbb{C}$ or $\mathbb{Z}_2$. The coefficients c^i should come from whatever our base field is (usually $\mathbb{R}$).

Example Consider the following vectors in $\mathbb{R}^3$:

$$v_1 = \begin{pmatrix}0\\0\\1\end{pmatrix}, \qquad v_2 = \begin{pmatrix}1\\2\\1\end{pmatrix}, \qquad v_3 = \begin{pmatrix}1\\2\\3\end{pmatrix}.$$

Are they linearly independent?

We need to see whether the system

$$c^1 v_1 + c^2 v_2 + c^3 v_3 = 0$$

has any solutions for c^1, c^2, c^3. We can rewrite this as a homogeneous system:

$$\begin{pmatrix}v_1 & v_2 & v_3\end{pmatrix}\begin{pmatrix}c^1\\c^2\\c^3\end{pmatrix} = 0.$$

This system has solutions if and only if the matrix $M = \begin{pmatrix}v_1 & v_2 & v_3\end{pmatrix}$ is singular, so we should find the determinant of M:

$$\det M = \det \begin{pmatrix}0 & 1 & 1\\0 & 2 & 2\\1 & 1 & 3\end{pmatrix} = \det \begin{pmatrix}1 & 1\\2 & 2\end{pmatrix} = 0.$$

Therefore nontrivial solutions exist. At this point we know that the vectors are linearly dependent. If we need to, we can find coefficients that demonstrate linear dependence by solving the system of equations:

$$\left(\begin{array}{ccc|c}0 & 1 & 1 & 0\\0 & 2 & 2 & 0\\1 & 1 & 3 & 0\end{array}\right) \sim \left(\begin{array}{ccc|c}1 & 1 & 3 & 0\\0 & 1 & 1 & 0\\0 & 0 & 0 & 0\end{array}\right) \sim \left(\begin{array}{ccc|c}1 & 0 & 2 & 0\\0 & 1 & 1 & 0\\0 & 0 & 0 & 0\end{array}\right).$$

Then $c^3 = \mu$, $c^2 = -\mu$, and $c^1 = -2\mu$. Now any choice of μ will produce coefficients c^1, c^2, c^3 that satisfy the linear equation. So we can set $\mu = 1$ and obtain:

$$c^1 v_1 + c^2 v_2 + c^3 v_3 = 0 \Rightarrow -2v_1 - v_2 + v_3 = 0.$$

Reading homework: problem 16.1

Theorem 16.1 (Linear Dependence)**.** *A set of non-zero vectors $\{v_1, \ldots, v_n\}$ is linearly dependent if and only if one of the vectors v_k is expressible as a linear combination of the preceeding vectors.*

Proof. The theorem is an if and only if statement, so there are two things to show.

i. First, we show that if $v_k = c^1 v_1 + \cdots c^{k-1} v_{k-1}$ then the set is linearly dependent.

This is easy. We just rewrite the assumption:

$$c^1 v_1 + \cdots + c^{k-1} v_{k-1} - v_k + 0 v_{k+1} + \cdots + 0 v_n = 0.$$

This is a vanishing linear combination of the vectors $\{v_1, \ldots, v_n\}$ with not all coefficients equal to zero, so $\{v_1, \ldots, v_n\}$ is a linearly dependent set.

ii. Now, we show that linear dependence implies that there exists k for which v_k is a linear combination of the vectors $\{v_1, \ldots, v_{k-1}\}$.

The assumption says that

$$c^1 v_1 + c^2 v_2 + \cdots + c^n v_n = 0.$$

Take k to be the largest number for which c_k is not equal to zero. So:

$$c^1 v_1 + c^2 v_2 + \cdots + c^{k-1} v_{k-1} + c^k v_k = 0.$$

(Note that $k > 1$, since otherwise we would have $c^1 v_1 = 0 \Rightarrow v_1 = 0$, contradicting the assumption that none of the v_i are the zero vector.)

As such, we can rearrange the equation:

$$\begin{array}{rcl} c^1 v_1 + c^2 v_2 + \cdots + c^{k-1} v_{k-1} & = & -c^k v_k \\ \Rightarrow \ -\frac{c^1}{c^k} v_1 - \frac{c^2}{c^k} v_2 - \cdots - \frac{c^{k-1}}{c^k} v_{k-1} & = & v_k. \end{array}$$

Therefore we have expressed v_k as a linear combination of the previous vectors, and we are done.

$\square$

Example Consider the vector space $P_2(t)$ of polynomials of degree less than or equal to 2. Set:

$$\begin{aligned} v_1 &= 1+t \\ v_2 &= 1+t^2 \\ v_3 &= t+t^2 \\ v_4 &= 2+t+t^2 \\ v_5 &= 1+t+t^2. \end{aligned}$$

The set $\{v_1, \ldots, v_5\}$ is linearly dependent, because $v_4 = v_1 + v_2$.

We have seen two different ways to show a set of vectors is linearly dependent: we can either find a linear combination of the vectors which is equal to zero, or we can express one of the vectors as a linear combination of the other vectors. On the other hand, to check that a set of vectors is linearly *independent*, we must check that no matter every non-zero linear combination of our vectors gives something other than the zero vector. Equivalently, to show that the set $v_1, v_2, \ldots, v_n$ is linearly independent, we must show that the equation $c_1v_1 + c_2v_2 + \cdots + c_nv_n = 0$ has no solutions other than $c_1 = c_2 = \cdots = c_n = 0$.

Example Consider the following vectors in $\mathbb{R}^3$:

$$v_1 = \begin{pmatrix}0\\0\\2\end{pmatrix}, \qquad v_2 = \begin{pmatrix}2\\2\\1\end{pmatrix}, \qquad v_3 = \begin{pmatrix}1\\4\\3\end{pmatrix}.$$

Are they linearly independent?

We need to see whether the system

$$c^1v_1 + c^2v_2 + c^3v_3 = 0$$

has any solutions for c^1, c^2, c^3. We can rewrite this as a homogeneous system:

$$\begin{pmatrix}v_1 & v_2 & v_3\end{pmatrix}\begin{pmatrix}c^1\\c^2\\c^3\end{pmatrix} = 0.$$

This system has solutions if and only if the matrix $M = \begin{pmatrix}v_1 & v_2 & v_3\end{pmatrix}$ is singular, so we should find the determinant of M:

$$\det M = \det\begin{pmatrix}0 & 2 & 1\\0 & 2 & 4\\2 & 1 & 3\end{pmatrix} = 2\det\begin{pmatrix}2 & 1\\2 & 4\end{pmatrix} = 12.$$

Since the matrix M has non-zero determinant, the only solution to the system of equations

$$\begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{pmatrix} c^1 \\ c^2 \\ c^3 \end{pmatrix} = 0$$

is $c_1 = c_2 = c_3 = 0$. (Why?) So the vectors v_1, v_2, v_3 are linearly independent.

Reading homework: problem 16.2

Now suppose vectors $v_1, \ldots, v_n$ are linearly dependent,

$$c^1 v_1 + c^2 v_2 + \cdots + c^n v_n = 0$$

with $c^1 \neq 0$. Then:

$$\operatorname{span}\{v_1, \ldots, v_n\} = \operatorname{span}\{v_2, \ldots, v_n\}$$

because any $x \in \operatorname{span}\{v_1, \ldots, v_n\}$ is given by

$$\begin{aligned} x &= a^1 v_1 + \cdots a^n v_n \\ &= a^1 \left(-\frac{c^2}{c_1} v_2 - \cdots - \frac{c^n}{c_1} v_n \right) + a^2 v_2 + \cdots + a^n v_n \\ &= \left(a^2 - a^1 \frac{c^2}{c_1} \right) v_2 + \cdots + \left(a^n - a^1 \frac{c^n}{c_1} \right) v_n. \end{aligned}$$

Then x is in $\operatorname{span}\{v_2, \ldots, v_n\}$.

When we write a vector space as the span of a list of vectors, we would like that list to be as short as possible (we will explore this idea further in lecture 17). This can be achieved by iterating the above procedure.

Example In the above example, we found that $v_4 = v_1 + v_2$. In this case, any expression for a vector as a linear combination involving v_4 can be turned into a combination without v_4 by making the substitution $v_4 = v_1 + v_2$.

Then:

$$\begin{aligned} S &= \operatorname{span}\{1+t, 1+t^2, t+t^2, 2+t+t^2, 1+t+t^2\} \\ &= \operatorname{span}\{1+t, 1+t^2, t+t^2, 1+t+t^2\}. \end{aligned}$$

Now we notice that $1+t+t^2 = \frac{1}{2}(1+t) + \frac{1}{2}(1+t^2) + \frac{1}{2}(t+t^2)$. So the vector $1+t+t^2 = v_5$ is also extraneous, since it can be expressed as a linear combination of the remaining three vectors, v_1, v_2, v_3. Therefore

$$S = \operatorname{span}\{1+t, 1+t^2, t+t^2\}.$$

In fact, you can check that there are no (non-zero) solutions to the linear system

$$c^1(1+t)+c^2(1+t^2)+c^3(t+t^2)=0.$$

Therefore the remaining vectors $\{1+t, 1+t^2, t+t^2\}$ are linearly independent, and span the vector space S. Then these vectors are a minimal spanning set, in the sense that no more vectors can be removed since the vectors are linearly independent. Such a set is called a *basis* for S.

Example Let B^3 be the space of 3×1 bit-valued matrices (i.e., column vectors). Is the following subset linearly independent?

$$\left\{ \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix} \right\}$$

If the set is linearly dependent, then we can find non-zero solutions to the system:

$$c^1\begin{pmatrix}1\\1\\0\end{pmatrix}+c^2\begin{pmatrix}1\\0\\1\end{pmatrix}+c^3\begin{pmatrix}0\\1\\1\end{pmatrix}=0,$$

which becomes the linear system

$$\begin{pmatrix}1&1&0\\1&0&1\\0&1&1\end{pmatrix}\begin{pmatrix}c^1\\c^2\\c^3\end{pmatrix}=0.$$

Solutions exist if and only if the determinant of the matrix is non-zero. But:

$$\det\begin{pmatrix}1&1&0\\1&0&1\\0&1&1\end{pmatrix}=1\det\begin{pmatrix}0&1\\1&1\end{pmatrix}-1\det\begin{pmatrix}1&1\\0&1\end{pmatrix}=-1-1=1+1=0$$

Therefore non-trivial solutions exist, and the set is not linearly independent.

References

Hefferon, Chapter Two, Section II: Linear Independence
Hefferon, Chapter Two, Section III.1: Basis
Beezer, Chapter V, Section LI
Beezer, Chapter V, Section LDS
Beezer, Chapter VS, Section LISS, Subsection LI
Wikipedia:

- Linear Independence
- Basis

Review Questions

1. Let B^n be the space of $n \times 1$ bit-valued matrices (*i.e.*, column vectors) over the field $\mathbb{Z}/2$. Remember that this means that the coefficients in any linear combination can be only 0 or 1, with rules for adding and multiplying coefficients given here.

   (a) How many different vectors are there in B^n?

   (b) Find a collection S of vectors that span B^3 and are linearly independent. In other words, find a basis of B^3.

   (c) Write each other vector in B^3 as a linear combination of the vectors in the set S that you chose.

   (d) Would it be possible to span B^3 with only two vectors?

2. Let e_i be the vector in $\mathbb{R}^n$ with a 1 in the ith position and 0's in every other position. Let v be an arbitrary vector in $\mathbb{R}^n$.

   (a) Show that the collection $\{e_1, \ldots, e_n\}$ is linearly independent.

   (b) Demonstrate that $v = \sum_{i=1}^{n}(v \cdot e_i)e_i$.

   (c) The span$\{e_1, \ldots, e_n\}$ is the same as what vector space?

17 Basis and Dimension

In Lecture 16, we established the notion of a linearly independent set of vectors in a vector space V, and of a set of vectors that span V. We saw that any set of vectors that span V can be reduced to some minimal collection of linearly independent vectors; such a set is called a *basis* of the subspace V.

Definition Let V be a vector space. Then a set S is a *basis* for V if S is linearly independent and $V = \operatorname{span} S$.

If S is a basis of V and S has only finitely many elements, then we say that V is *finite-dimensional*. The number of vectors in S is the *dimension* of V.

Suppose V is a *finite-dimensional* vector space, and S and T are two different bases for V. One might worry that S and T have a different number of vectors; then we would have to talk about the dimension of V in terms of the basis S or in terms of the basis T. Luckily this isn't what happens. Later in this section, we will show that S and T must have the same number of vectors. This means that the dimension of a vector space does not depend on the basis. In fact, dimension is a very important way to characterize of any vector space V.

Example $P_n(t)$ has a basis $\{1, t, \ldots, t^n\}$, since every polynomial of degree less than or equal to n is a sum

$$a^0\, 1 + a^1\, t + \cdots + a^n\, t^n, \qquad a^i \in \mathbb{R}$$

so $P_n(t) = \operatorname{span}\{1, t, \ldots, t^n\}$. This set of vectors is linearly independent: If the polynomial $p(t) = c^0 1 + c^1 t + \cdots + c^n t^n = 0$, then $c^0 = c^1 = \cdots = c^n = 0$, so $p(t)$ is the zero polynomial.

Then $P_n(t)$ is finite dimensional, and $\dim P_n(t) = n + 1$.

Theorem 17.1. *Let $S = \{v_1, \ldots, v_n\}$ be a basis for a vector space V. Then every vector $w \in V$ can be written* uniquely *as a linear combination of vectors in the basis S:*

$$w = c^1 v_1 + \cdots + c^n v_n.$$

Proof. Since S is a basis for V, then $\operatorname{span} S = V$, and so there exist constants c^i such that $w = c^1 v_1 + \cdots + c^n v_n$.

Suppose there exists a second set of constants d^i such that

$$w = d^1 v_1 + \cdots + d^n v_n \,.$$

Then:

$$\begin{aligned} 0_V &= w - w \\ &= c^1 v_1 + \cdots + c^n v_n - d^1 v_1 + \cdots + d^n v_n \\ &= (c^1 - d^1) v_1 + \cdots + (c^n - d^n) v_n. \end{aligned}$$

If it occurs exactly once that $c^i \neq d^i$, then the equation reduces to $0 = (c^i - d^i) v_i$, which is a contradiction since the vectors v_i are assumed to be non-zero.

If we have more than one i for which $c^i \neq d^i$, we can use this last equation to write one of the vectors in S as a linear combination of other vectors in S, which contradicts the assumption that S is linearly independent. Then for every i, $c^i = d^i$. $\square$

Next, we would like to establish a method for determining whether a collection of vectors forms a basis for $\mathbb{R}^n$. But first, we need to show that any two bases for a finite-dimensional vector space has the same number of vectors.

Lemma 17.2. *If $S = \{v_1, \dots, v_n\}$ is a basis for a vector space V and $T = \{w_1, \dots, w_m\}$ is a linearly independent set of vectors in V, then $m \leq n$.*

The idea of the proof is to start with the set S and replace vectors in S one at a time with vectors from T, such that after each replacement we still have a basis for V.

Reading homework: problem 17.1

Proof. Since S spans V, then the set $\{w_1, v_1, \dots, v_n\}$ is linearly dependent. Then we can write w_1 as a linear combination of the v_i; using that equation, we can express one of the v_i in terms of w_1 and the remaining v_j with $j \neq i$. Then we can discard one of the v_i from this set to obtain a linearly independent set that still spans V. Now we need to prove that S_1 is a basis; we need to show that S_1 is linearly independent and that S_1 spans V.

The set $S_1 = \{w_1, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n\}$ is linearly independent: By the previous theorem, there was a unique way to express w_1 in terms of the set S. Now, to obtain a contradiction, suppose there is some k and constants c^i such that

$$v_k = c^0 w_1 + c^1 v_1 + \cdots + c^{i-1} v_{i-1} + c^{i+1} v_{i+1} + \cdots + c^n v_n.$$

Then replacing w_1 with its expression in terms of the collection S gives a way to express the vector v_k as a linear combination of the vectors in S, which contradicts the linear independence of S. On the other hand, we cannot express w_1 as a linear combination of the vectors in $\{v_j | j \neq i\}$, since the expression of w_1 in terms of S was unique, and had a non-zero coefficient on the vector v_i. Then no vector in S_1 can be expressed as a combination of other vectors in S_1, which demonstrates that S_1 is linearly independent.

The set S_1 spans V: For any $u \in V$, we can express u as a linear combination of vectors in S. But we can express v_i as a linear combination of vectors in the collection S_1; rewriting v_i as such allows us to express u as a linear combination of the vectors in S_1.

Then S_1 is a basis of V with n vectors.

We can now iterate this process, replacing one of the v_i in S_1 with w_2, and so on. If $m \leq n$, this process ends with the set $S_m = \{w_1, \ldots, w_m, v_{i_1}, \ldots, v_{i_{n-m}}\}$, which is fine.

Otherwise, we have $m > n$, and the set $S_n = \{w_1, \ldots, w_n\}$ is a basis for V. But we still have some vector w_{n+1} in T that is not in S_n. Since S_n is a basis, we can write w_{n+1} as a combination of the vectors in S_n, which contradicts the linear independence of the set T. Then it must be the case that $m \leq n$, as desired. □

Corollary 17.3. *For a finite-dimensional vector space V, any two bases for V have the same number of vectors.*

Proof. Let S and T be two bases for V. Then both are linearly independent sets that span V. Suppose S has n vectors and T has m vectors. Then by the previous lemma, we have that $m \leq n$. But (exchanging the roles of S and T in application of the lemma) we also see that $n \leq m$. Then $m = n$, as desired. □

Reading homework: problem 17.2

17.1 Bases in $\mathbb{R}^n$.

From one of the review questions, we know that

$$\mathbb{R}^n = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\},$$

and that this set of vectors is linearly independent. So this set of vectors is a basis for $\mathbb{R}^n$, and $\dim \mathbb{R}^n = n$. This basis is often called the *standard* or *canonical basis* for $\mathbb{R}^n$. The vector with a one in the ith position and zeros everywhere else is written e_i. It points in the direction of the ith coordinate axis, and has unit length. In multivariable calculus classes, this basis is often written $\{i, j, k\}$ for $\mathbb{R}^3$.

Bases are not unique. While there exists a unique way to express a vector in terms of any particular basis, bases themselves are far from unique. For example, both of the sets:

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \text{ and } \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

are bases for $\mathbb{R}^2$. Rescaling any vector in one of these sets is already enough to show that $\mathbb{R}^2$ has infinitely many bases. But even if we require that all of the basis vectors have unit length, it turns out that there are still infinitely many bases for $\mathbb{R}^2$. (See Review Question 3.)

To see whether a collection of vectors $S = \{v_1, \ldots, v_m\}$ is a basis for $\mathbb{R}^n$, we have to check that they are linearly independent and that they span $\mathbb{R}^n$. From the previous discussion, we also know that m must equal n, so assume S has n vectors.

If S is linearly independent, then there is no non-trivial solution of the equation

$$0 = x^1 v_1 + \cdots + x^n v_n.$$

Let M be a matrix whose columns are the vectors v_i. Then the above equation is equivalent to requiring that there is a unique solution to

$$MX = 0\,.$$

To see if S spans $\mathbb{R}^n$, we take an arbitrary vector w and solve the linear system
$$w = x^1 v_1 + \cdots + x^n v_n$$
in the unknowns c^i. For this, we need to find a unique solution for the linear system $MX = w$.

Thus, we need to show that M^{-1} exists, so that
$$X = M^{-1} w$$
is the unique solution we desire. Then we see that S is a basis for V if and only if $\det M \neq 0$.

Theorem 17.4. *Let $S = \{v_1, \ldots, v_m\}$ be a collection of vectors in $\mathbb{R}^n$. Let M be the matrix whose columns are the vectors in S. Then S is a basis for V if and only if m is the dimension of V and*
$$\det M \neq 0.$$

Example Let
$$S = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \text{ and } T = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}.$$

Then set $M_S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Since $\det M_S = 1 \neq 0$, then S is a basis for $\mathbb{R}^2$.

Likewise, set $M_T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. Since $\det M_T = -2 \neq 0$, then T is a basis for $\mathbb{R}^2$.

References

Hefferon, Chapter Two, Section II: Linear Independence
Hefferon, Chapter Two, Section III.1: Basis
Beezer, Chapter VS, Section B, Subsections B-BNM
Beezer, Chapter VS, Section D, Subsections D-DVS
Wikipedia:

- Linear Independence
- Basis

Review Questions

1. (a) Draw the collection of all unit vectors in $\mathbb{R}^2$.

   (b) Let $S_x = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, x \right\}$, where x is a unit vector in $\mathbb{R}^2$. For which x is S_x a basis of $\mathbb{R}^2$?

2. Let B^n be the vector space of column vectors with bit entries $0, 1$. Write down every basis for B^1 and B^2. How many bases are there for B^3? B^4? Can you make a conjecture for the number of bases for B^n?

   (Hint: You can build up a basis for B^n by choosing one vector at a time, such that the vector you choose is not in the span of the previous vectors you've chosen. How many vectors are in the span of any one vector? Any two vectors? How many vectors are in the span of any k vectors, for $k \leq n$?)

3. Suppose that V is an n-dimensional vector space.

   (a) Show that any n linearly independent vectors in V form a basis. (Hint: Let $\{w_1, \ldots, w_m\}$ be a collection of n linearly independent vectors in V, and let $\{v_1, \ldots, v_n\}$ be a basis for V. Apply the method of Lemma 19.2 to these two sets of vectors.)

   (b) Show that any set of n vectors in V which span V forms a basis for V.

   (Hint: Suppose that you have a set of n vectors which span V but do not form a basis. What must be true about them? How could you get a basis from this set? Use Corollary 19.3 to derive a contradiction.)

4. Let S be a collection of vectors $\{v_1, \ldots, v_n\}$ in a vector space V. Show that if every vector w in V can be expressed uniquely as a linear combination of vectors in S, then S is a basis of V. In other words: suppose that for every vector w in V, there is exactly one set of constants $c^1, \ldots, c^n$ so that $c^1 v_1 + \cdots + c^n v_n = w$. Show that this means that the set S is linearly independent and spans V. (This is the converse to the theorem in the lecture.)

5. Vectors are objects that you can add together; show that the set of all linear transformations mapping $\mathbb{R}^3 \to \mathbb{R}$ is itself a vector space. Find a

basis for this vector space. Do you think your proof could be modified to work for linear transformations $\mathbb{R}^n \to \mathbb{R}$?

(Hint: Represent $\mathbb{R}^3$ as column vectors, and argue that a linear transformation $T\colon \mathbb{R}^3 \to \mathbb{R}$ is just a row vector.)

(Hint: If you are stuck or just curious, look up "dual space.")

18 Eigenvalues and Eigenvectors

Matrix of a Linear Transformation Consider a linear transformation

$$L\colon \mathbb{R}^2 \to \mathbb{R}^2 \,.$$

Suppose we know that $L\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}a\\c\end{pmatrix}$ and $L\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}b\\d\end{pmatrix}$. Then, because of linearity, we can determine what L does to any vector $\begin{pmatrix}x\\y\end{pmatrix}$:

$$L\begin{pmatrix}x\\y\end{pmatrix} = L(x\begin{pmatrix}1\\0\end{pmatrix}+y\begin{pmatrix}0\\1\end{pmatrix}) = xL\begin{pmatrix}1\\0\end{pmatrix}+yL\begin{pmatrix}0\\1\end{pmatrix} = x\begin{pmatrix}a\\c\end{pmatrix}+y\begin{pmatrix}b\\d\end{pmatrix} = \begin{pmatrix}ax+by\\cx+dy\end{pmatrix}.$$

Now notice that for any vector $\begin{pmatrix}x\\y\end{pmatrix}$, we have

$$\begin{pmatrix}a & b\\c & d\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}ax+by\\cx+dy\end{pmatrix} = L\begin{pmatrix}x\\y\end{pmatrix}.$$

Then the matrix $\begin{pmatrix}a & b\\c & d\end{pmatrix}$ acts by matrix multiplication in the same way that L does. Call this matrix the *matrix of* L in the *"basis"* $\left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\right\}$.

Since every linear function from $\mathbb{R}^2 \to \mathbb{R}^2$ can be given a matrix in this way, we see that every such linear function has a matrix in the basis $\left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\right\}$. We will revisit this idea in depth later, and develop the notion of a basis further, and learn about how to make a matrix for an arbitrary linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ in an arbitrary basis.

18.1 Invariant Directions

Consider the linear transformation L such that

$$L\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}-4\\-10\end{pmatrix} \text{ and } L\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}3\\7\end{pmatrix},$$

so that the matrix of L is $\begin{pmatrix}-4 & 3\\-10 & 7\end{pmatrix}$. Recall that a vector is a direction and a magnitude; L applied to $\begin{pmatrix}1\\0\end{pmatrix}$ or $\begin{pmatrix}0\\1\end{pmatrix}$ changes both the direction and the magnitude of the vectors given to it.

Notice that $L\begin{pmatrix}3\\5\end{pmatrix} = \begin{pmatrix}-4\cdot 3+3\cdot 5\\-10\cdot 3+7\cdot 5\end{pmatrix} = \begin{pmatrix}3\\5\end{pmatrix}$. Then L fixes both the magnitude and direction of the vector $v_1 = \begin{pmatrix}3\\5\end{pmatrix}$. *Try drawing a picture of this situation on some graph paper to help yourself visualize it better!*

Reading homework: problem 18.1

Now, notice that any vector with the same direction as v_1 can be written as cv_1 for some constant c. Then $L(cv_1) = cL(v_1) = cv_1$, so L fixes every vector pointing in the same direction as v_1.

Also notice that $L\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}-4\cdot 1+3\cdot 2\\-10\cdot 1+7\cdot 2\end{pmatrix} = \begin{pmatrix}2\\4\end{pmatrix} = 2\begin{pmatrix}1\\2\end{pmatrix}$. Then L fixes the direction of the vector $v_2 = \begin{pmatrix}1\\2\end{pmatrix}$ but stretches v_2 by a factor of 2. Now notice that for any constant c, $L(cv_2) = cL(v_2) = 2cv_2$. Then L stretches every vector pointing in the same direction as v_2 by a factor of 2.

In short, given a linear transformation L it is sometimes possible to find a vector $v \neq 0$ and constant $\lambda \neq 0$ such that

$$L(v) = \lambda v.$$

We call the direction of the vector v an *invariant direction*. In fact, any vector pointing in the same direction also satisfies the equation: $L(cv) = cL(v) = \lambda cv$. The vector v is called an *eigenvector* of L, and λ is an *eigenvalue*. Since the direction is all we really care about here, then any other vector cv (so long as $c \neq 0$) is an equally good choice of eigenvector. Notice that the relation "u and v point in the same direction" is an equivalence relation.

Returning to our example of the linear transformation L with matrix $\begin{pmatrix}-4 & 3\\-10 & 7\end{pmatrix}$, we have seen that L enjoys the property of having two invariant directions, represented by eigenvectors v_1 and v_2 with eigenvalues 1 and 2, respectively.

It would be very convenient if I could write any vector w as a linear combination of v_1 and v_2. Suppose $w = rv_1 + sv_2$ for some constants r and s. Then:

$$L(w) = L(rv_1 + sv_2) = rL(v_1) + sL(v_2) = rv_1 + 2sv_2.$$

Now L just multiplies the number r by 1 and the number s by 2. If we could write this as a matrix, it would look like:

$$\begin{pmatrix}1 & 0\\ 0 & 2\end{pmatrix}\begin{pmatrix}r\\ s\end{pmatrix}$$

which is much slicker than the usual scenario $L\begin{pmatrix}x\\ y\end{pmatrix}=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=\begin{pmatrix}ax+by\\ cx+dy\end{pmatrix}$. Here, r and s give the coordinates of w in terms of the vectors v_1 and v_2. In the previous example, we multiplied the vector by the matrix L and came up with a complicated expression. In these coordinates, we can see that L is a very simple *diagonal matrix*, whose diagonal entries are exactly the *eigenvalues* of L.

This process is called *diagonalization*, and it can make complicated linear systems much easier to analyze.

Reading homework: problem 18.2

Now that we've seen what eigenvalues and eigenvectors are, there are a number of questions that need to be answered.

- How do we find eigenvectors and their eigenvalues?
- How many eigenvalues and (independent) eigenvectors does a given linear transformation have?
- When can a linear transformation be diagonalized?

We'll start by trying to find the eigenvectors for a linear transformation.

Example Let $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $L(x,y)=(2x+2y,16x+6y)$. First, we can find the matrix of L:

$$\begin{pmatrix}x\\ y\end{pmatrix} \stackrel{L}{\longmapsto} \begin{pmatrix}2 & 2\\ 16 & 6\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}.$$

We want to find an invariant direction $v=\begin{pmatrix}x\\ y\end{pmatrix}$ such that

$$L(v)=\lambda v$$

or, in matrix notation,

$$\begin{aligned}
\begin{pmatrix} 2 & 2 \\ 16 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \lambda \begin{pmatrix} x \\ y \end{pmatrix} \\
\Leftrightarrow \begin{pmatrix} 2 & 2 \\ 16 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
\Leftrightarrow \begin{pmatrix} 2-\lambda & 2 \\ 16 & 6-\lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
\end{aligned}$$

This is a homogeneous system, so it only has solutions when the matrix $\begin{pmatrix} 2-\lambda & 2 \\ 16 & 6-\lambda \end{pmatrix}$ is singular. In other words,

$$\begin{aligned}
\det \begin{pmatrix} 2-\lambda & 2 \\ 16 & 6-\lambda \end{pmatrix} &= 0 \\
\Leftrightarrow (2-\lambda)(6-\lambda) - 32 &= 0 \\
\Leftrightarrow \lambda^2 - 8\lambda - 20 &= 0 \\
\Leftrightarrow (\lambda - 10)(\lambda + 2) &= 0
\end{aligned}$$

For any square $n \times n$ matrix M, the polynomial in λ given by

$$P_M(\lambda) = \det(\lambda I - M) = (-1)^n \det(M - \lambda I)$$

is called the *characteristic polynomial* of M, and its roots are the eigenvalues of M.

In this case, we see that L has two eigenvalues, $\lambda_1 = 10$ and $\lambda_2 = -2$. To find the eigenvectors, we need to deal with these two cases separately. To do so, we solve the linear system $\begin{pmatrix} 2-\lambda & 2 \\ 16 & 6-\lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ with the particular eigenvalue λ plugged in to the matrix.

$\underline{\lambda = 10}$: We solve the linear system

$$\begin{pmatrix} -8 & 2 \\ 16 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Both equations say that $y = 4x$, so any vector $\begin{pmatrix} x \\ 4x \end{pmatrix}$ will do. Since we only need the direction of the eigenvector, we can pick a value for x. Setting $x = 1$ is convenient, and gives the eigenvector $v_1 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$.

$\underline{\lambda = -2}$: We solve the linear system

$$\begin{pmatrix} 4 & 2 \\ 16 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Here again both equations agree, because we chose λ to make the system singular. We see that $y = -2x$ works, so we can choose $v_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$.

In short, our process was the following:

- Find the characteristic polynomial of the matrix M for L, given by[8] $\det(\lambda I - M)$.
- Find the roots of the characteristic polynomial; these are the eigenvalues of L.
- For each eigenvalue λ_i, solve the linear system $(M - \lambda_i I)v = 0$ to obtain an eigenvector v associated to λ_i.

References

Hefferon, Chapter Three, Section III.1: Representing Linear Maps with Matrices
Hefferon, Chapter Five, Section II.3: Eigenvalues and Eigenvectors
Beezer, Chapter E, Section EE
Wikipedia:

- Eigen*
- Characteristic Polynomial
- Linear Transformations (and matrices thereof)

Review Questions

1. Let $M = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$. Find all eigenvalues of M. Does M have two independent[9] eigenvectors? Can M be diagonalized?
2. Consider $L \colon \mathbb{R}^2 \to \mathbb{R}^2$ with $L(x, y) = (x \cos\theta + y \sin\theta, -x \sin\theta + y \cos\theta)$.

[8]It is often easier (and equivalent if you only need the roots) to compute $\det(M - \lambda I)$.
[9]Independence of vectors will be explained later; for now, think "not parallel".

(a) Write the matrix of L in the basis $\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}$.

(b) When $\theta \neq 0$, explain how L acts on the plane. Draw a picture.

(c) Do you expect L to have invariant directions?

(d) Try to find real eigenvalues for L by solving the equation

$$L(v) = \lambda v.$$

(e) Are there complex eigenvalues for L, assuming that $i = \sqrt{-1}$ exists?

3. Let L be the linear transformation $L\colon \mathbb{R}^3 \to \mathbb{R}^3$ given by $L(x, y, z) = (x + y, x + z, y + z)$. Let e_i be the vector with a one in the ith position and zeros in all other positions.

(a) Find Le_i for each i.

(b) Given a matrix $M = \begin{pmatrix} m^1_1 & m^1_2 & m^1_3 \\ m^2_1 & m^2_2 & m^2_3 \\ m^3_1 & m^3_2 & m^3_3 \end{pmatrix}$, what can you say about Me_i for each i?

(c) Find a 3×3 matrix M representing L. Choose three nonzero vectors pointing in different directions and show that $Mv = Lv$ for each of your choices.

(d) Find the eigenvectors and eigenvalues of M.

19 Eigenvalues and Eigenvectors II

In Lecture 18, we developed the idea of eigenvalues and eigenvectors in the case of linear transformations $\mathbb{R}^2 \to \mathbb{R}^2$. In this section, we will develop the idea more generally.

Definition For a linear transformation $L\colon V \to V$, then λ is an *eigenvalue* of L with *eigenvector* $v \neq 0_V$ if

$$Lv = \lambda v.$$

This equation says that the direction of v is invariant (unchanged) under L.

Let's try to understand this equation better in terms of matrices. Let V be a finite-dimensional vector space (we'll explain what it means to be finite-dimensional in more detail later; for now, take this to mean $\mathbb{R}^n$), and let $L\colon V \to V$.

Matrix of a Linear Transformation Any vector in $\mathbb{R}^n$ can be written as a linear combination of the *standard basis vectors* $\{e_i | i \in \{1, \ldots, n\}\}$. The vector e_i has a one in the ith position, and zeros everywhere else. *I.e.*

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Then to find the matrix of any linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^n$, it suffices to know what $L(e_i)$ is for every i.

For any matrix M, observe that Me_i is equal to the ith column of M. Then if the ith column of M equals $L(e_i)$ for every i, then $Mv = L(v)$ for every $v \in \mathbb{R}^n$. Then the matrix representing L in the standard basis is just the matrix whose ith column is $L(e_i)$.

Since we can represent L by a square matrix M, we find eigenvalues λ and associated eigenvectors v by solving the homogeneous system

$$(M - \lambda I)v = 0.$$

This system has non-zero solutions if and only if the matrix

$$M - \lambda I$$

is singular, and so we require that

$$\det(\lambda I - M) = 0.$$

The left hand side of this equation is a polynomial in the variable λ called the *characteristic polynomial* $P_M(\lambda)$ of M. For an $n \times n$ matrix, the characteristic polynomial has degree n. Then

$$P_M(\lambda) = \lambda^n + c_1 \lambda^{n-1} + \cdots + c_n.$$

Notice that $P_M(0) = \det(-M) = (-1)^n \det M$.

The *fundamental theorem of algebra* states that any polynomial can be factored into a product of linear terms over $\mathbb{C}$. Then there exists a collection of n complex numbers λ_i (possibly with repetition) such that

$$P_M(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n), \qquad P_M(\lambda_i) = 0$$

The eigenvalues λ_i of M are exactly the roots of $P_M(\lambda)$. These eigenvalues could be real or complex or zero, and they need not all be different. The number of times that any given root λ_i appears in the collection of eigenvalues is called its *multiplicity*.

Example Let L be the linear transformation $L \colon \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$L(x, y, z) = (2x + y - z, x + 2y - z, -x - y + 2z).$$

The matrix M representing L has columns Le_i for each i, so:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \stackrel{L}{\mapsto} \begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Then the characteristic polynomial of L is[10]

$$\begin{aligned} P_M(\lambda) &= \det \begin{pmatrix} \lambda - 2 & -1 & 1 \\ -1 & \lambda - 2 & 1 \\ 1 & 1 & \lambda - 2 \end{pmatrix} \\ &= (\lambda - 2)[(\lambda - 2)^2 - 1] + [-(\lambda - 2) - 1] + [-(\lambda - 2) - 1] \\ &= (\lambda - 1)^2 (\lambda - 4) \end{aligned}$$

[10] It is often easier (and equivalent) to solve $\det(M - \lambda I) = 0$.

Then L has eigenvalues $\lambda_1 = 1$ (with multiplicity 2), and $\lambda_2 = 4$ (with multiplicity 1).

To find the eigenvectors associated to each eigenvalue, we solve the homogeneous system $(M - \lambda_i I)X = 0$ for each i.

$\underline{\lambda = 4}$: We set up the augmented matrix for the linear system:

$$\left(\begin{array}{ccc|c} -2 & 1 & -1 & 0 \\ 1 & -2 & -1 & 0 \\ -1 & -1 & -2 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -2 & -1 & 0 \\ 0 & -3 & -3 & 0 \\ 0 & -3 & -3 & 0 \end{array}\right)$$
$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

So we see that $z = t$, $y = -t$, and $x = -t$ gives a formula for eigenvectors in terms of the free parameter t. Any such eigenvector is of the form $t \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$; thus L leaves a line through the origin invariant.

$\underline{\lambda = 1}$: Again we set up an augmented matrix and find the solution set:

$$\left(\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 1 & 1 & -1 & 0 \\ -1 & -1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Then the solution set has two free parameters, s and t, such that $z = t$, $y = s$, and $x = -s + t$. Then L leaves invariant the set:

$$\left\{ s \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \middle| s, t \in \mathbb{R} \right\}.$$

This set is a plane through the origin. So the multiplicity two eigenvalue has two independent eigenvectors, $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ that determine an invariant plane.

Example Let V be the vector space of smooth (*i.e.* infinitely differentiable) functions $f \colon \mathbb{R} \to \mathbb{R}$. Then the derivative is a linear operator $\frac{d}{dx} \colon V \to V$. What are the eigenvectors of the derivative? In this case, we don't have a matrix to work with, so we have to make do.

A function f is an eigenvector of $\frac{d}{dx}$ if there exists some number λ such that $\frac{d}{dx}f = \lambda f$. An obvious candidate is the exponential function, $e^{\lambda x}$; indeed, $\frac{d}{dx}e^{\lambda x} = \lambda e^{\lambda x}$.

As such, the operator $\frac{d}{dx}$ has an eigenvector $e^{\lambda x}$ for every $\lambda \in \mathbb{R}$.

This is actually the whole collection of eigenvectors for $\frac{d}{dx}$; this can be proved using the fact that every infinitely differentiable function has a Taylor series with infinite radius of convergence, and then using the Taylor series to show that if two functions are eigenvectors of $\frac{d}{dx}$ with eigenvalues λ, then they are scalar multiples of each other.

19.1 Eigenspaces

In the previous example, we found two eigenvectors $\begin{pmatrix}-1\\1\\0\end{pmatrix}$ and $\begin{pmatrix}1\\0\\1\end{pmatrix}$ for L with eigenvalue 1. Notice that $\begin{pmatrix}-1\\1\\0\end{pmatrix}+\begin{pmatrix}1\\0\\1\end{pmatrix}=\begin{pmatrix}0\\1\\1\end{pmatrix}$ is also an eigenvector of L with eigenvalue 1. In fact, any linear combination $r\begin{pmatrix}-1\\1\\0\end{pmatrix}+s\begin{pmatrix}1\\0\\1\end{pmatrix}$ of these two eigenvectors will be another eigenvector with the same eigenvalue.

More generally, let $\{v_1, v_2, \ldots\}$ be eigenvectors of some linear transformation L with the same eigenvalue λ. A *linear combination* of the v_i can be written $c_1v_1 + c_2v_2 + \cdots$ for some constants $\{c_1, c_2, \ldots\}$. Then:

$$\begin{aligned} L(c_1v_1 + c_2v_2 + \cdots) &= c_1Lv_1 + c_2Lv_2 + \cdots \text{ by linearity of } L \\ &= c_1\lambda v_1 + c_2\lambda v_2 + \cdots \text{ since } Lv_i = \lambda v_i \\ &= \lambda(c_1v_1 + c_2v_2 + \cdots). \end{aligned}$$

So every linear combination of the v_i is an eigenvector of L with the same eigenvalue λ. In simple terms, any sum of eigenvectors is again an eigenvector *if they share the same eigenvalue.*

The space of all vectors with eigenvalue λ is called an *eigenspace*. It is, in fact, a vector space contained within the larger vector space V: It contains 0_V, since $L0_V = 0_V = \lambda 0_V$, and is closed under addition and scalar multiplication by the above calculation. All other vector space properties are inherited from the fact that V itself is a vector space.

An eigenspace is an example of a *subspace* of V, a notion that we will explore further next time.

Reading homework: problem 19.1

You are now ready to attempt the second sample midterm.

References

Hefferon, Chapter Three, Section III.1: Representing Linear Maps with Matrices
Hefferon, Chapter Five, Section II.3: Eigenvalues and Eigenvectors
Beezer, Chapter E, Section EE
Wikipedia:

- Eigen*
- Characteristic Polynomial
- Linear Transformations (and matrices thereof)

Review Questions

1. Explain why the characteristic polynomial of an $n \times n$ matrix has degree n. Make your explanation easy to read by starting with some simple examples, and then use properties of the determinant to give a *general* explanation.

2. Compute the characteristic polynomial $P_M(\lambda)$ of the matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Now, since we can evaluate polynomials on square matrices, we can plug M into its characteristic polynomial and find the *matrix* $P_M(M)$. What do you find from this computation? Does something similar hold for 3×3 matrices? What about $n \times n$ matrices?

3. *Discrete dynamical system.* Let M be the matrix given by

$$M = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}.$$

Given any vector $v(0) = \begin{pmatrix} x(0) \\ y(0) \end{pmatrix}$, we can create an infinite sequence of vectors $v(1), v(2), v(3)$, and so on using the rule

$$v(t+1) = Mv(t) \text{ for all natural numbers } t.$$

(This is known as a *discrete dynamical system* whose *initial condition* is $v(0)$.)

(a) Find all eigenvectors and eigenvalues of M.

(b) Find all vectors $v(0)$ such that

$$v(0) = v(1) = v(2) = v(3) = \cdots$$

(Such a vector is known as a *fixed point* of the dynamical system.)

(c) Find all vectors $v(0)$ such that $v(0), v(1), v(2), v(3), \ldots$ all point in the same direction. (Any such vector describes an *invariant curve* of the dynamical system.)

20 Diagonalization

Given a linear transformation, we are interested in how to write it as a matrix. We are especially interested in the case that the matrix is written with respect to a basis of eigenvectors, in which case it is a particularly nice matrix. But first, we discuss matrices of linear transformations.

20.1 Matrix of a Linear Transformation

Let V and W be vector spaces, with bases $S = \{e_1, \ldots, e_n\}$ and $T = \{f_1, \ldots, f_m\}$ respectively. Since these are bases, there exist constants v^i and w^j such that any vectors $v \in V$ and $w \in W$ can be written as:

$$\begin{aligned} v &= v^1 e_1 + v^2 e_2 + \cdots + v^n e_n \\ w &= w^1 f_1 + w^2 f_2 + \cdots + w^m f_m \end{aligned}$$

We call the coefficients $v^1, \ldots, v^n$ the *components* of v in the basis[11] $\{e_1, \ldots, e_n\}$. It is often convenient to arrange the components v^i in a column vector and the basis vector in a row vector by writing

$$v = \begin{pmatrix} e_1 & e_2 & \cdots & e_n \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix} .$$

Example Consider the basis $S = \{1 - t, 1 + t\}$ for the vector space $P_1(t)$. The vector $v = 2t$ has components $v^1 = -1, v^2 = 1$, because

$$v = -1(1 - t) + 1(1 + t) = \begin{pmatrix} 1 - t & 1 + t \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} .$$

We may consider these components as vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$:

$$\begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \in \mathbb{R}^n, \qquad \begin{pmatrix} w^1 \\ \vdots \\ w^m \end{pmatrix} \in \mathbb{R}^m.$$

[11] To avoid confusion, it helps to notice that components of a vector are almost always labeled by a superscript, while basis vectors are labeled by subscripts in the conventions of these lecture notes.

Now suppose we have a linear transformation $L\colon V \to W$. Then we can expect to write L as an $m \times n$ matrix, turning an n-dimensional vector of coefficients corresponding to v into an m-dimensional vector of coefficients for w.

Using linearity, we write:

$$\begin{aligned} L(v) &= L(v^1 e_1 + v^2 e_2 + \cdots + v^n e_n) \\ &= v^1 L(e_1) + v^2 L(e_2) + \cdots + v^n L(e_n) \\ &= \begin{pmatrix} L(e_1) & L(e_2) & \cdots & L(e_n) \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix}. \end{aligned}$$

This is a vector in W. Let's compute its components in W.

We know that for each e_j, $L(e_j)$ is a vector in W, and can thus be written uniquely as a linear combination of vectors in the basis T. Then we can find coefficients M^i_j such that:

$$L(e_j) = f_1 M^1_j + \cdots + f_m M^m_j = \sum_{i=1}^{m} f_i M^i_j = \begin{pmatrix} f_1 & f_2 & \cdots & f_m \end{pmatrix} \begin{pmatrix} M^1_j \\ M^2_j \\ \vdots \\ M^m_j \end{pmatrix}.$$

We've written the M^i_j on the right side of the f's to agree with our previous notation for matrix multiplication. We have an "up-hill rule" where the matching indices for the multiplied objects run up and to the right, like so: $f_i M^i_j$.

Now M^i_j is the ith component of $L(e_j)$. Regarding the coefficients M^i_j as a matrix, we can see that the jth column of M is the coefficients of $L(e_j)$ in the basis T.

Then we can write:

$$
\begin{aligned}
L(v) &= L(v^1e_1+v^2e_2+\cdots+v^ne_n) \\
&= v^1L(e_1)+v^2L(e_2)+\cdots+v^nL(e_n) \\
&= \sum_{i=1}^{m} L(e_j)v^j \\
&= \sum_{i=1}^{m} (M_j^1 f_1+\cdots+M_j^m f_m)v^j \\
&= \sum_{i=1}^{m} f_i \left[\sum_{j=1}^{n} M_j^i v^j\right] \\
&= \begin{pmatrix} f_1 & f_2 & \cdots & f_m \end{pmatrix} \begin{pmatrix} M_1^1 & M_2^1 & \cdots & M_n^1 \\ M_1^2 & M_2^2 & & \\ \vdots & & & \vdots \\ M_1^m & & \cdots & M_n^m \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix}
\end{aligned}
$$

The second last equality is the definition of matrix multiplication which is obvious from the last line. Thus:

$$
\begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \stackrel{L}{\mapsto} \begin{pmatrix} M_1^1 & \dots & M_n^1 \\ \vdots & & \vdots \\ M_1^m & \dots & M_n^m \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix},
$$

and $M = (M_j^i)$ is called the matrix of L. Notice that this matrix depends on a *choice* of bases for both V and W. Also observe that the columns of M are computed by examining L acting on each basis vector in V expanded in the basis vectors of W.

Example Let $L\colon P_1(t) \mapsto P_1(t)$, such that $L(a+bt) = (a+b)t$. Since $V = P_1(t) = W$, let's choose the same basis for V and W. We'll choose the basis $\{1-t, 1+t\}$ for this example.

Thus:

$$
\begin{aligned}
L(1-t) &= (1-1)t = \ 0 \ = (1-t)\cdot 0 + (1+t)\cdot 0 = \begin{pmatrix}(1-t) & (1+t)\end{pmatrix}\begin{pmatrix}0\\0\end{pmatrix} \\
L(1+t) &= (1+1)t = 2t \ = (1-t)\cdot -1 + (1+t)\cdot 1 = \begin{pmatrix}(1-t) & (1+t)\end{pmatrix}\begin{pmatrix}-1\\1\end{pmatrix} \\
\Rightarrow M &= \begin{pmatrix}0 & -1\\0 & 1\end{pmatrix}
\end{aligned}
$$

To obtain the last line we used that fact that the columns of M are just the coefficients of L on each of the basis vectors; this always makes it easy to write down M in terms of the basis we have chosen.

Reading homework: problem 20.1

20.2 Diagonalization

Now suppose we are lucky, and we have $L\colon V \mapsto V$, and the basis $\{v_1, \ldots, v_n\}$ is a set of linearly independent eigenvectors for L, with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then:

$$\begin{array}{rcl} L(v_1) & = & \lambda_1 v_1 \\ L(v_2) & = & \lambda_2 v_2 \\ & \vdots & \\ L(v_n) & = & \lambda_n v_n \end{array}$$

As a result, the matrix of L in the basis of eigenvectors is:

$$M = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix},$$

where all entries off of the diagonal are zero.

Suppose that V is any n-dimensional vector space. We call a linear transformation $L\colon V \mapsto V$ *diagonalizable* if there exists a collection of n linearly independent eigenvectors for L. In other words, L is diagonalizable if there exists a basis for V of eigenvectors for L.

In a basis of eigenvectors, the matrix of a linear transformation is diagonal. On the other hand, if an $n \times n$ matrix M is diagonal, then the standard basis vectors e_i are already a set of n linearly independent eigenvectors for M. We have shown:

Theorem 20.1. *Given a basis S for a vector space V and a linear transformation $L\colon V \to V$, then the matrix for L in the basis S is diagonal if and only if S is a basis of eigenvectors for L.*

Reading homework: problem 20.2

20.3 Change of Basis

Suppose we have two bases $S = \{v_1, \ldots, v_n\}$ and $T = \{u_1, \ldots, u_n\}$ for a vector space V. (Here v_i and u_i are *vectors*, not components of vectors in a basis!) Then we may write each v_i uniquely as a linear combination of the u_j:

$$v_j = \sum_i u_i P^i_j \,,$$

or in a matrix notation

$$\begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \begin{pmatrix} P^1_1 & P^1_2 & \cdots & P^1_n \\ P^2_1 & P^2_2 & & \\ \vdots & & & \vdots \\ P^n_1 & & \cdots & P^n_n \end{pmatrix}.$$

Here, the P^i_j are constants, which we can regard as entries of a square matrix $P = (P^i_j)$. The matrix P must have an inverse, since we can also write each u_i uniquely as a linear combination of the v_j:

$$u_j = \sum_k v_k Q^k_j.$$

Then we can write:

$$v_j = \sum_k \sum_i v_k Q^k_j P^i_j.$$

But $\sum_i Q^k_j P^i_j$ is the k, j entry of the product of the matrices QP. Since the only expression for v_j in the basis S is v_j itself, then QP fixes each v_j. As a result, each v_j is an eigenvector for QP with eigenvalues 1, so QP is the identity.

The matrix P is called a *change of basis* matrix.

Changing basis changes the matrix of a linear transformation. However, as a map between vector spaces, the linear transformation is the same no matter which basis we use. Linear transformations are the actual objects of study of this course, not matrices; matrices are merely a convenient way of doing computations.

To wit, suppose $L \colon V \mapsto V$ has matrix $M = (M^i_j)$ in the basis $T = \{u_1, \ldots, u_n\}$, so

$$L(u_i) = \sum_k M^k_i u_k.$$

Now, let $S = \{v_1, \ldots, v_n\}$ be a basis of eigenvectors for L, with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

$$L(v_i) = \lambda_i v_i = \sum_k v_k D_i^k$$

where D is the diagonal matrix whose diagonal entries D_k^k are the eigenvalues λ_k; ie, $D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$. Let P be the change of basis matrix from the basis T to the basis S. Then:

$$L(v_j) = L\left(\sum_i u_i P_j^i\right) = \sum_i L(u_i) P_j^i = \sum_i \sum_k u_k M_i^k P_j^i.$$

Meanwhile, we have:

$$L(v_i) = \sum_k v_k D_i^k = \sum_k \sum_j u_j P_k^j D_i^k.$$

Since the expression for a vector in a basis is unique, then we see that the entries of MP are the same as the entries of PD. In other words, we see that

$$MP = PD \qquad \text{or} \qquad D = P^{-1}MP.$$

This motivates the following definition:

Definition A matrix M is *diagonalizable* if there exists an invertible matrix P and a diagonal matrix D such that

$$D = P^{-1}MP.$$

We can summarize as follows:

- Change of basis multiplies vectors by the change of basis matrix P, to give vectors in the new basis.
- To get the matrix of a linear transformation in the new basis, we *conjugate* the matrix of L by the change of basis matrix: $M \to P^{-1}MP$.

If for two matrices N and M there exists an invertible matrix P such that $M = P^{-1}NP$, then we say that M and N are *similar*. Then the above discussion shows that diagonalizable matrices are similar to diagonal matrices.

Corollary 20.2. *A square matrix M is diagonalizable if and only if there exists a basis of eigenvectors for M. Moreover, these eigenvectors are the columns of the change of basis matrix P which diagonalizes M.*

Reading homework: problem 20.3

Example Let's try to diagonalize the matrix

$$M = \begin{pmatrix} -14 & -28 & -44 \\ -7 & -14 & -23 \\ 9 & 18 & 29 \end{pmatrix}.$$

The eigenvalues of M are determined by

$$\det(M - \lambda) = -\lambda^3 + \lambda^2 + 2\lambda = 0.$$

So the eigenvalues of M are $-1, 0,$ and 2, and associated eigenvectors turn out to be $v_1 = \begin{pmatrix} -8 \\ -1 \\ 3 \end{pmatrix}, v_2 = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$, and $v_3 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$. In order for M to be diagonalizable, we need the vectors v_1, v_2, v_3 to be linearly independent. Notice that the matrix

$$P = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} = \begin{pmatrix} -8 & -2 & -1 \\ -1 & 1 & -1 \\ 3 & 0 & 1 \end{pmatrix}$$

is invertible because its determinant is -1. Therefore, the eigenvectors of M form a basis of $\mathbb{R}$, and so M is diagonalizable. Moreover, the matrix P of eigenvectors is a change of basis matrix which diagonalizes M:

$$P^{-1}MP = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

References

Hefferon, Chapter Three, Section V: Change of Basis
Beezer, Chapter E, Section SD
Beezer, Chapter R, Sections MR-CB
Wikipedia:

- Change of Basis
- Diagonalizable Matrix
- Similar Matrix

Review Questions

1. Let $P_n(t)$ be the vector space of polynomials of degree n or less, and $\frac{d}{dt}\colon P_n(t) \mapsto P_{n-1}(t)$ be the derivative operator. Find the matrix of $\frac{d}{dt}$ in the bases $\{1, t, \ldots, t^n\}$ for $P_n(t)$ and $\{1, t, \ldots, t^{n-1}\}$ for $P_{n-1}(t)$.

2. When writing a matrix for a linear transformation, we have seen that the choice of basis matters. In fact, even the order of the basis matters!

    - Write all possible reorderings of the standard basis $\{e_1, e_2, e_3\}$ for $\mathbb{R}^3$.
    - Write each change of basis matrix between the standard basis $\{e_1, e_2, e_3\}$ and each of its reorderings. Make as many observations as you can about these matrices: what are their entries? Do you notice anything about how many of each type of entry appears in each row and column? What are their determinants? (Note: These matrices are known as *permutation matrices*.)
    - Given the linear transformation $L(x, y, z) = (2y-z, 3x, 2z+x+y)$, write the matrix M for L in the standard basis, and two other reorderings of the standard basis. How are these matrices related?

3. When is the 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ diagonalizable? Include examples in your answer.

4. Show that similarity of matrices is an *equivalence relation*. (The definition of an equivalence relation is given in Homework 0.)

5. *Jordan form*

    - Can the matrix $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$ be diagonalized? Either diagonalize it or explain why this is impossible.

- Can the matrix $\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$ be diagonalized? Either diagonalize it or explain why this is impossible.
- Can the $n \times n$ matrix $\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}$ be diagonalized? Either diagonalize it or explain why this is impossible.

  Note: It turns out that every complex matrix is similar to a block matrix whose diagonal blocks look like diagonal matrices or the ones above and whose off-diagonal blocks are all zero. This is called the *Jordan form* of the matrix.

21 Orthonormal Bases

The canonical/standard basis

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

has many useful properties.

- Each of the standard basis vectors has unit length:

$$\|e_i\| = \sqrt{e_i \cdot e_i} = \sqrt{e_i^T e_i} = 1.$$

- The standard basis vectors are *orthogonal* (in other words, at right angles or perpendicular).

$$e_i \cdot e_j = e_i^T e_j = 0 \text{ when } i \neq j$$

This is summarized by

$$e_i^T e_j = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases},$$

where δ_{ij} is the *Kronecker delta*. Notice that the Kronecker delta gives the entries of the identity matrix.

Given column vectors v and w, we have seen that the dot product $v \cdot w$ is the same as the matrix multiplication $v^T w$. This is the *inner product* on $\mathbb{R}^n$. We can also form the *outer product* vw^T, which gives a square matrix.

The outer product on the standard basis vectors is interesting. Set

$$\begin{aligned}
\Pi_1 &= e_1 e_1^T \\
&= \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \\
&\vdots \\
\Pi_n &= e_n e_n^T \\
&= \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & \cdots & 1 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}
\end{aligned}$$

In short, Π_i is the diagonal square matrix with a 1 in the ith diagonal position and zeros everywhere else. [12]

Notice that $\Pi_i \Pi_j = e_i e_i^T e_j e_j^T = e_i \delta_{ij} e_j^T$. Then:

$$\Pi_i \Pi_j = \left\{ \begin{array}{cc} \Pi_i & i = j \\ 0 & i \neq j \end{array} \right. .$$

Moreover, for a diagonal matrix D with diagonal entries $\lambda_1, \ldots, \lambda_n$, we can write

$$D = \lambda_1 \Pi_1 + \cdots + \lambda_n \Pi_n.$$

[12] This is reminiscent of an older notation, where vectors are written in juxtaposition. This is called a "dyadic tensor", and is still used in some applications.

Other bases that share these properties should behave in many of the same ways as the standard basis. As such, we will study:

- *Orthogonal bases* $\{v_1, \ldots, v_n\}$:

$$v_i \cdot v_j = 0 \text{ if } i \neq j$$

  In other words, all vectors in the basis are perpendicular.

- *Orthonormal bases* $\{u_1, \ldots, u_n\}$:

$$u_i \cdot u_j = \delta_{ij}.$$

  In addition to being orthogonal, each vector has unit length.

Suppose $T = \{u_1, \ldots, u_n\}$ is an orthonormal basis for $\mathbb{R}^n$. Since T is a basis, we can write any vector v uniquely as a linear combination of the vectors in T:

$$v = c^1 u_1 + \cdots c^n u_n.$$

Since T is orthonormal, there is a very easy way to find the coefficients of this linear combination. By taking the dot product of v with any of the vectors in T, we get:

$$\begin{array}{rcl}
v \cdot u_i & = & c^1 u_1 \cdot u_i + \cdots + c^i u_i \cdot u_i + \cdots + c^n u_n \cdot u_i \\
& = & c^1 \cdot 0 + \cdots + c^i \cdot 1 + \cdots + c^n \cdot 0 \\
& = & c^i, \\
\Rightarrow c^i & = & v \cdot u_i \\
\Rightarrow v & = & (v \cdot u_1) u_1 + \cdots + (v \cdot u_n) u_n \\
& = & \sum_i (v \cdot u_i) u_i.
\end{array}$$

This proves the theorem:

Theorem 21.1. *For an orthonormal basis $\{u_1, \ldots, u_n\}$, any vector v can be expressed as*

$$v = \sum_i (v \cdot u_i) u_i.$$

Reading homework: problem 21.1

21.1 Relating Orthonormal Bases

Suppose $T = \{u_1, \ldots, u_n\}$ and $R = \{w_1, \ldots, w_n\}$ are two orthonormal bases for $\mathbb{R}^n$. Then:

$$\begin{aligned}
w_1 &= (w_1 \cdot u_1)u_1 + \cdots + (w_1 \cdot u_n)u_n \\
&\vdots \\
w_n &= (w_n \cdot u_1)u_1 + \cdots + (w_n \cdot u_n)u_n \\
\Rightarrow w_i &= \sum_j u_j (u_j \cdot w_i)
\end{aligned}$$

As such, the matrix for the change of basis from T to R is given by

$$P = (P_i^j) = (u_j \cdot w_i).$$

Consider the product PP^T in this case.

$$\begin{aligned}
(PP^T)_k^j &= \sum_i (u_j \cdot w_i)(w_i \cdot u_k) \\
&= \sum_i (u_j^T w_i)(w_i^T u_k) \\
&= u_j^T \left[\sum_i (w_i w_i^T) \right] u_k \\
&= u_j^T I_n u_k \qquad (*) \\
&= u_j^T u_k = \delta_{jk}.
\end{aligned}$$

The equality $(*)$ is explained below. So assuming $(*)$ holds, we have shown that $PP^T = I_n$, which implies that

$$P^T = P^{-1}.$$

The equality in the line $(*)$ says that $\sum_i w_i w_i^T = I_n$. To see this, we examine $\left(\sum_i w_i w_i^T\right) v$ for an arbitrary vector v. We can find constants c^j

such that $v = \sum_j c^j w_j$, so that:

$$\begin{aligned}
\left(\sum_i w_i w_i^T\right) v &= \left(\sum_i w_i w_i^T\right)\left(\sum_j c^j w_j\right) \\
&= \sum_j c^j \sum_i w_i w_i^T w_j \\
&= \sum_j c^j \sum_i w_i \delta_{ij} \\
&= \sum_j c^j w_j \text{ since all terms with } i \neq j \text{ vanish} \\
&= v.
\end{aligned}$$

Then as a linear transformation, $\sum_i w_i w_i^T = I_n$ fixes every vector, and thus must be the identity I_n.

Definition A matrix P is *orthogonal* if $P^{-1} = P^T$.

Then to summarize,

Theorem 21.2. *A change of basis matrix P relating two orthonormal bases is an orthogonal matrix. I.e.,*

$$P^{-1} = P^T.$$

Reading homework: problem 21.2

Example Consider $\mathbb{R}^3$ with the orthonormal basis

$$S = \left\{ u_1 = \begin{pmatrix} \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \end{pmatrix}, u_2 = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, u_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix} \right\}.$$

Let R be the standard basis $\{e_1, e_2, e_3\}$. Since we are changing from the standard basis to a new basis, then the columns of the change of basis matrix are exactly the images of the standard basis vectors. Then the change of basis matrix from R to S is given by:

$$
\begin{aligned}
P = (P_i^j) = (e_j u_i) &= \begin{pmatrix} e_1 \cdot u_1 & e_1 \cdot u_2 & e_1 \cdot u_3 \\ e_2 \cdot u_1 & e_2 \cdot u_2 & e_2 \cdot u_3 \\ e_3 \cdot u_1 & e_3 \cdot u_2 & e_3 \cdot u_3 \end{pmatrix} \\
= \begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix} &= \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{3}} \\ \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix}.
\end{aligned}
$$

From our theorem, we observe that:

$$
\begin{aligned}
P^{-1} = P^T &= \begin{pmatrix} u_1^T \\ u_2^T \\ u_3^T \end{pmatrix} \\
&= \begin{pmatrix} \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}.
\end{aligned}
$$

We can check that $P^T P = I$ by a lengthy computation, or more simply, notice that

$$
\begin{aligned}
(P^T P)_{ij} &= \begin{pmatrix} u_1^T \\ u_2^T \\ u_3^T \end{pmatrix} \begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\end{aligned}
$$

We are using orthonormality of the u_i for the matrix multiplication above. It is very important to realize that the columns of an *orthogonal* matrix are made from an *orthonormal* set of vectors.

Orthonormal Change of Basis and Diagonal Matrices. Suppose D is a diagonal matrix, and we use an orthogonal matrix P to change to a new basis. Then the matrix M of D in the new basis is:

$$M = PDP^{-1} = PDP^T.$$

Now we calculate the transpose of M.

$$\begin{aligned} M^T &= (PDP^T)^T \\ &= (P^T)^T D^T P^T \\ &= PDP^T \\ &= M \end{aligned}$$

So we see the matrix PDP^T is symmetric!

References

Hefferon, Chapter Three, Section V: Change of Basis
Beezer, Chapter V, Section O, Subsection N
Beezer, Chapter VS, Section B, Subsection OBC
Wikipedia:

- Orthogonal Matrix
- Diagonalizable Matrix
- Similar Matrix

Review Questions

1. Let $D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$.

   (a) Write D in terms of the vectors e_1 and e_2, and their transposes.

   (b) Suppose $P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible. Show that D is similar to

   $$M = \frac{1}{ad - bc} \begin{pmatrix} \lambda_1 ad - \lambda_2 bc & -(\lambda_1 - \lambda_2)ab \\ (\lambda_1 - \lambda_2)cd & -\lambda_1 bc + \lambda_2 ad \end{pmatrix}.$$

   (c) Suppose the vectors $\begin{pmatrix} a & b \end{pmatrix}$ and $\begin{pmatrix} c & d \end{pmatrix}$ are orthogonal. What can you say about M in this case? (Hint: think about what M^T is equal to.)

2. Suppose $S = \{v_1, \ldots, v_n\}$ is an *orthogonal* (not orthonormal) basis for $\mathbb{R}^n$. Then we can write any vector v as $v = \sum_i c^i v_i$ for some constants c^i. Find a formula for the constants c^i in terms of v and the vectors in S.

3. Let u, v be independent vectors in $\mathbb{R}^3$, and $P = \operatorname{span}\{u, v\}$ be the plane spanned by u and v.

   (a) Is the vector $v^\perp = v - \frac{u \cdot v}{u \cdot u} u$ in the plane P?

   (b) What is the angle between $v^\perp$ and u?

   (c) Given your solution to the above, how can you find a third vector perpendicular to both u and $v^\perp$?

   (d) Construct an orthonormal basis for $\mathbb{R}^3$ from u and v.

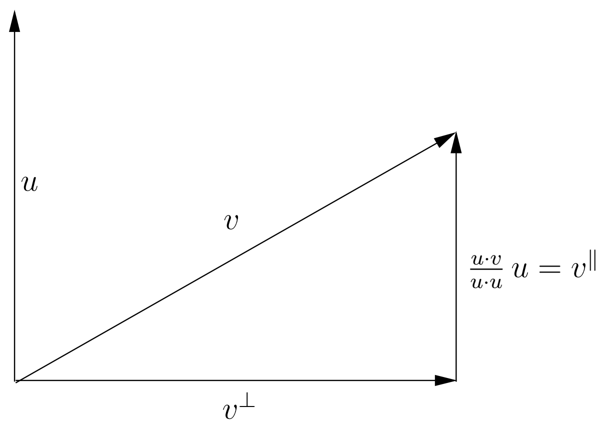
   (e) Test your abstract formulae starting with

   $$u = \begin{pmatrix} 1 & 2 & 0 \end{pmatrix} \text{ and } v = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}.$$

22 Gram-Schmidt and Orthogonal Complements

Given a vector u and some other vector v not in the span of u, we can construct a new vector:

$$v^{\perp} = v - \frac{u \cdot v}{u \cdot u} u.$$



This new vector $v^{\perp}$ is orthogonal to u because

$$u \cdot v^{\perp} = u \cdot v - \frac{u \cdot v}{u \cdot u} u \cdot u = 0.$$

Hence, $\{u, v^{\perp}\}$ is an orthogonal basis for $\operatorname{span}\{u, v\}$. When v is not parallel to u, $v^{\perp} \neq 0$, and normalizing these vectors we obtain $\left\{ \frac{u}{|u|}, \frac{v^{\perp}}{|v^{\perp}|} \right\}$, an orthonormal basis.

Sometimes we write $v = v^{\perp} + v^{\parallel}$ where:

$$\begin{aligned} v^{\perp} &= v - \frac{u \cdot v}{u \cdot u} u \\ v^{\parallel} &= \frac{u \cdot v}{u \cdot u} u. \end{aligned}$$

This is called an *orthogonal decomposition* because we have decomposed v into a sum of orthogonal vectors. It is significant that we wrote this decomposition with u in mind; $v^{\parallel}$ is parallel to u.

If u, v are linearly independent vectors in $\mathbb{R}^3$, then the set $\{u, v^{\perp}, u \times v^{\perp}\}$ would be an orthogonal basis for $\mathbb{R}^3$. This set could then be normalized by dividing each vector by its length to obtain an orthonormal basis.

However, it often occurs that we are interested in vector spaces with dimension greater than 3, and must resort to craftier means than cross products to obtain an orthogonal basis. [13]

Given a third vector w, we should first check that w does not lie in the span of u and v, *i.e.* check that u, v and w are linearly independent. We then can define:

$$w^{\perp} = w - \frac{u \cdot w}{u \cdot u}\, u - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}}\, v^{\perp}.$$

We can check that $u \cdot w^{\perp}$ and $v^{\perp} \cdot w^{\perp}$ are both zero:

$$\begin{aligned} u \cdot w^{\perp} &= u \cdot \left(w - \frac{u \cdot w}{u \cdot u}\, u - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}}\, v^{\perp} \right) \\ &= u \cdot w - \frac{u \cdot w}{u \cdot u} u \cdot u - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}} u \cdot v^{\perp} \\ &= u \cdot w - u \cdot w - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}} u \cdot v^{\perp} \ = \ 0 \end{aligned}$$

since u is orthogonal to $v^{\perp}$, and

$$\begin{aligned} v^{\perp} \cdot w^{\perp} &= v^{\perp} \cdot \left(w - \frac{u \cdot w}{u \cdot u}\, u - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}}\, v^{\perp} \right) \\ &= v^{\perp} \cdot w - \frac{u \cdot w}{u \cdot u} v^{\perp} \cdot u - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}} v^{\perp} \cdot v^{\perp} \\ &= v^{\perp} \cdot w - \frac{u \cdot w}{u \cdot u} v^{\perp} \cdot u - v^{\perp} \cdot w \ = \ 0 \end{aligned}$$

because u is orthogonal to $v^{\perp}$. Since $w^{\perp}$ is orthogonal to both u and $v^{\perp}$, we have that $\{u, v^{\perp}, w^{\perp}\}$ is an orthogonal basis for $\operatorname{span}\{u, v, w\}$.

In fact, given a collection $\{x, v_2, \ldots\}$ of linearly independent vectors, we can produce an orthogonal basis for $\operatorname{span}\{v_1, v_2, \ldots\}$ consisting of the follow-

[13] Actually, given a set T of $(n-1)$ independent vectors in n-space, one can define an analogue of the cross product that will produce a vector orthogonal to the span of T, using a method exactly analogous to the usual computation for calculating the cross product of two vectors in $\mathbb{R}^3$. This only gets us the *last* orthogonal vector, though; the process in this Section gives a way to get a full orthogonal basis.

ing vectors:

$$\begin{aligned}
v_1^\perp &= v_1 \\
v_2^\perp &= v_2 - \frac{v_1^\perp \cdot v_2}{v_1^\perp \cdot v_1^\perp}\, v_1^\perp \\
v_3^\perp &= v_3 - \frac{v_1^\perp \cdot v_3}{v_1^\perp \cdot v_1^\perp}\, v_1^\perp - \frac{v_2^\perp \cdot v_3}{v_2^\perp \cdot v_2^\perp}\, v_2^\perp \\
&\vdots \\
v_i^\perp &= v_i - \sum_{j<i} \frac{v_j^\perp \cdot v_i}{v_j^\perp \cdot v_j^\perp}\, v_j^\perp \\
&= v_i - \frac{v_1^\perp \cdot v_i}{v_1^\perp \cdot v_1^\perp}\, v_1^\perp - \cdots - \frac{v_{n-1}^\perp \cdot v_i}{v_{n-1}^\perp \cdot v_{n-1}^\perp}\, v_{n-1}^\perp \\
&\vdots
\end{aligned}$$

Notice that each $v_i^\perp$ here depends on the existence of $v_j^\perp$ for every $j < i$. This allows us to inductively/algorithmically build up a linearly independent, orthogonal set of vectors whose span is $\operatorname{span}\{v_1, v_2, \ldots\}$. This algorithm bears the name *Gram–Schmidt orthogonalization procedure*.

Example Let $u = \begin{pmatrix}1 & 1 & 0\end{pmatrix}, v = \begin{pmatrix}1 & 1 & 1\end{pmatrix}$, and $w = \begin{pmatrix}3 & 1 & 1\end{pmatrix}$. We'll apply Gram-Schmidt to obtain an orthogonal basis for $\mathbb{R}^3$.

First, we set $u^\perp = u$. Then:

$$\begin{aligned}
v^\perp &= \begin{pmatrix}1 & 1 & 1\end{pmatrix} - \frac{2}{2}\begin{pmatrix}1 & 1 & 0\end{pmatrix} = \begin{pmatrix}0 & 0 & 1\end{pmatrix} \\
w^\perp &= \begin{pmatrix}3 & 1 & 1\end{pmatrix} - \frac{4}{2}\begin{pmatrix}1 & 1 & 0\end{pmatrix} - \frac{1}{1}\begin{pmatrix}0 & 0 & 1\end{pmatrix} = \begin{pmatrix}1 & -1 & 0\end{pmatrix}.
\end{aligned}$$

Then the set

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

is an orthogonal basis for $\mathbb{R}^3$. To obtain an orthonormal basis, as always we simply divide each of these vectors by its length, yielding:

$$\left\{\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\end{pmatrix}, \begin{pmatrix}0 & 0 & 1\end{pmatrix}, \begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0\end{pmatrix}\right\}.$$

In Lecture 11 we learnt how to solve linear systems by decomposing a matrix M into a product of lower and upper triangular matrices

$$M = LU\,.$$

The Gram–Schmidt procedure suggests another matrix decomposition,

$$M = QR$$

where Q is an orthogonal matrix and R is an upper triangular matrix. So-called QR-decompositions are useful for solving linear systems, eigenvalue problems and least squares approximations. You can easily get the idea behind QR decomposition by working through a simple example.

Example Find the QR decomposition of

$$M = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 3 & -2 \\ 0 & 1 & -2 \end{pmatrix}\,.$$

What we will do is to think of the columns of M as three vectors and use Gram–Schmidt to build an orthonormal basis from these that will become the columns of the orthogonal matrix Q. We will use the matrix R to record the steps of the Gram–Schmidt procedure in such a way that the product QR equals M.

To begin with we write

$$M = \begin{pmatrix} 2 & -\frac{7}{5} & 1 \\ 1 & \frac{14}{5} & -2 \\ 0 & 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{5} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\,.$$

In the first matrix the first two columns are mutually orthogonal because we simpy replaced the second column of M by the vector that the Gram–Schmidt procedure produces from the first two columns of M, namely

$$\begin{pmatrix} -\frac{7}{5} \\ \frac{14}{5} \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix} - \frac{1}{5}\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}\,.$$

The matrix on the right is almost the identity matrix, save the $+\frac{1}{5}$ in the second entry of the first row, whose effect upon multiplying the two matrices precisely undoes what we we did to the second column of the first matrix.

For the third column of M we use Gram–Schmidt to deduce the third orthogonal vector

$$\begin{pmatrix}-\frac{1}{6}\\ \frac{1}{3}\\ -\frac{7}{6}\end{pmatrix}=\begin{pmatrix}1\\-2\\-2\end{pmatrix}-0.\begin{pmatrix}2\\1\\0\end{pmatrix}-\frac{-9}{\frac{54}{5}}\begin{pmatrix}-\frac{7}{5}\\ \frac{14}{5}\\ 1\end{pmatrix},$$

and therefore, using exactly the same procedure write

$$M=\begin{pmatrix}2 & -\frac{7}{5} & -\frac{1}{6}\\ 1 & \frac{14}{5} & \frac{1}{3}\\ 0 & 1 & -\frac{7}{6}\end{pmatrix}\begin{pmatrix}1 & \frac{1}{5} & 0\\ 0 & 1 & -\frac{5}{6}\\ 0 & 0 & 1\end{pmatrix}.$$

This is not quite the answer because the first matrix is now made of mutually orthogonal column vectors, but a *bona fide* orthogonal matrix is comprised of *orthonormal* vectors. To achieve that we divide each column of the first matrix by its length and multiply the corresponding row of the second matrix by the same amount:

$$M=\begin{pmatrix}\frac{2\sqrt{5}}{5} & -\frac{7\sqrt{30}}{90} & -\frac{\sqrt{6}}{18}\\ \frac{\sqrt{5}}{5} & \frac{7\sqrt{30}}{45} & \frac{\sqrt{6}}{9}\\ 0 & \frac{\sqrt{30}}{18} & -\frac{7\sqrt{6}}{18}\end{pmatrix}\begin{pmatrix}\sqrt{5} & \frac{\sqrt{5}}{5} & 0\\ 0 & \frac{3\sqrt{30}}{5} & -\frac{\sqrt{30}}{2}\\ 0 & 0 & \frac{\sqrt{6}}{2}\end{pmatrix}=QR\,.$$

A nice check of this result is to verify that entry (i,j) of the matrix R equals the dot product of the i-th column of Q with the j-th column of M. (Some people memorize this fact and use it as a recipe for computing QR deompositions.) *A good test of your own understanding is to work out why this is true!*

22.1 Orthogonal Complements

Let U and V be subspaces of a vector space W. We saw as a review exercise that $U\cap V$ is a subspace of W, and that $U\cup V$ was not a subspace. However, $\operatorname{span}(U\cup V)$ is certainly a subspace, since the span of *any* subset is a subspace.

Notice that all elements of $\operatorname{span}(U\cup V)$ take the form $u+v$ with $u\in U$ and $v\in V$. We call the subspace

$$U+V=\operatorname{span}(U\cup V)=\{u+v|u\in U,v\in V\}$$

the *sum* of U and V. Here, we are not adding vectors, but vector spaces to produce a new vector space!

Definition Given two subspaces U and V of a space W such that $U \cap V = \{0_W\}$, the *direct sum* of U and V is defined as:

$$U \oplus V = \operatorname{span}(U \cup V) = \{u + v | u \in U, v \in V\}.$$

Notice that when $U \cap V = \{0_W\}$, $U + V = U \oplus V$.

The direct sum has a very nice property.

Theorem 22.1. *Let $w = u + v \in U \oplus V$. Then the expression $w = u + v$ is unique. That is, there is only one way to write w as the sum of a vector in U and a vector in V.*

Proof. Suppose that $u + v = u' + v'$, with $u, u' \in U$, and $v, v' \in V$. Then we could express $0 = (u - u') + (v - v')$. Then $(u - u') = -(v - v')$. Since U and V are subspaces, we have $(u - u') \in U$ and $-(v - v') \in V$. But since these elements are equal, we also have $(u - u') \in V$. Since $U \cap V = \{0\}$, then $(u - u') = 0$. Similarly, $(v - v') = 0$, proving the theorem. □

Reading homework: problem 22.1

Given a subspace U in W, we would like to write W as the direct sum of U and *something*. Using the inner product, there is a natural candidate for this second subspace.

Definition Given a subspace U of a vector space W, define:

$$U^{\perp} = \{w \in W | w \cdot u = 0 \text{ for all } u \in U\}.$$

The set $U^{\perp}$ (pronounced "U-perp") is the set of all vectors in W orthogonal to *every* vector in U. This is also often called the *orthogonal complement* of U.

Example Consider any plane P through the origin in $\mathbb{R}^3$. Then P is a subspace, and $P^{\perp}$ is the line through the origin orthogonal to P. For example, if P is the xy-plane, then

$$\mathbb{R}^3 = P \oplus P^{\perp} = \{(x, y, 0) | x, y \in \mathbb{R}\} \oplus \{(0, 0, z) | z \in \mathbb{R}\}.$$

Theorem 22.2. *Let U be a subspace of a finite-dimensional vector space W. Then the set $U^{\perp}$ is a subspace of W, and $W = U \oplus U^{\perp}$.*

Proof. To see that $U^\perp$ is a subspace, we only need to check closure, which requires a simple check.

We have $U \cap U^\perp = \{0\}$, since if $u \in U$ and $u \in U^\perp$, we have:

$$u \cdot u = 0 \Leftrightarrow u = 0.$$

Finally, we show that any vector $w \in W$ is in $U \oplus U^\perp$. (This is where we use the assumption that W is finite-dimensional.) Let $e_1, \ldots, e_n$ be an orthonormal basis for W. Set:

$$\begin{aligned} u &= (w \cdot e_1)e_1 + \cdots + (w \cdot e_n)e_n \in U \\ u^\perp &= w - u \end{aligned}$$

It is easy to check that $u^\perp \in U^\perp$ (see the Gram-Schmidt procedure). Then $w = u + u^\perp$, so $w \in U \oplus U^\perp$, and we are done. □

Reading homework: problem 22.2

Example Consider any line L through the origin in $\mathbb{R}^4$. Then L is a subspace, and $L^\perp$ is a 3-dimensional subspace orthogonal to L. For example, let L be the line spanned by the vector $(1,1,1,1) \in \mathbb{R}^4$. Then Ł$^\perp$ is given by

$$\begin{aligned} L^\perp &= \{(x,y,z,w) \mid x,y,z,w \in \mathbb{R} \text{ and } (x,y,z,w) \cdot (1,1,1,1) = 0\} \\ &= \{(x,y,z,w) \mid x,y,z,w \in \mathbb{R} \text{ and } x,y,z,w = 0\}. \end{aligned}$$

It is easy to check that $\{v_1 = (1,-1,0,0), v_2 = (1,0,-1,0), v_3 = (1,0,0,-1)\}$ forms a basis for $L^\perp$. We use Gram-Schmidt to find an orthogonal basis for $L^\perp$:

First, we set $v_1^\perp = v_1$. Then:

$$\begin{aligned} v_2^\perp &= (1,0,-1,0) - \frac{1}{2}(1,-1,0,0) = \left(\frac{1}{2}, \frac{1}{2}, -1, 0\right), \\ v_3^\perp &= (1,0,0,-1) - \frac{1}{2}(1,-1,0,0) - \frac{1/2}{3/2}\left(\frac{1}{2}, \frac{1}{2}, -1, 0\right) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -1\right). \end{aligned}$$

So the set

$$\left\{(1,-1,0,0), \left(\frac{1}{2}, \frac{1}{2}, -1, 0\right), \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -1\right)\right\}$$

is an orthogonal basis for $L^\perp$. We find an orthonormal basis for $L^\perp$ by dividing each basis vector by its length:

$$\left\{\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, 0\right), \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}, 0\right), \left(\frac{\sqrt{3}}{6}, \frac{\sqrt{3}}{6}, \frac{\sqrt{3}}{6}, -\frac{\sqrt{3}}{2}\right)\right\}.$$

Moreover, we have

$$\mathbb{R}^4 = L \oplus L^\perp = \{(c,c,c,c) \mid c \in \mathbb{R}\} \oplus \{(x,y,z,w) \mid x,y,z,w \in \mathbb{R} \text{ and } x+y+z+w = 0\}.$$

Notice that for any subspace U, the subspace $(U^\perp)^\perp$ is just U again. As such, $\perp$ is an involution on the set of subspaces of a vector space.

References

Hefferon, Chapter Three, Section VI.2: Gram-Schmidt Orthogonalization
Beezer, Chapter V, Section O, Subsection GSP
Wikipedia:

- Gram-Schmidt Process
- QR Decomposition
- Orthonormal Basis
- Direct Sum

Review Questions

1. Find the QR factorization of
$$M = \begin{pmatrix} 1 & 0 & 2 \\ -1 & 2 & 0 \\ -1 & -2 & 2 \end{pmatrix}.$$

2. Suppose u and v are linearly independent. Show that u and $v^\perp$ are also linearly independent. Explain why $\{u, v^\perp\}$ are a basis for $\operatorname{span}\{u, v\}$.

3. Repeat the previous problem, but with three independent vectors u, v, w, and $v^\perp$ and $w^\perp$ as defined in the lecture.

4. Given any three vectors u, v, w, when do $v^\perp$ or $w^\perp$ vanish?

5. For U a subspace of W, use the subspace theorem to check that $U^\perp$ is a subspace of W.

6. This question will answer the question, “If I choose a bit vector *at random*, what is the probability that it lies in the span of some other vectors?”

   i. Given a collection S of k bit vectors in B^3, consider the bit matrix M whose columns are the vectors in S. Show that S is linearly independent if and only if the kernel of M is trivial.

   ii. Give some method for choosing a random bit vector v in B^3. Suppose S is a collection of 2 linearly independent bit vectors in B^3. How can we tell whether $S \cup \{v\}$ is linearly independent? Do you think it is likely or unlikely that $S \cup \{v\}$ is linearly independent? Explain your reasoning.

   iii. If P is the characteristic polynomial of a 3×3 bit matrix, what must the degree of P be? Given that each coefficient must be either 0 or 1, how many possibilities are there for P? How many of these possible characteristic polynomials have 0 as a root? If M is a 3×3 bit matrix chosen at random, what is the probability that it has 0 as an eigenvalue? (Assume that you are choosing a random matrix M in such a way as to make each characteristic polynomial equally likely.) What is the probability that the columns of M form a basis for B^3? (Hint: what is the relationship between the kernel of M and its eigenvalues?)

   Note: We could ask the same question for real vectors: If I choose a real vector at random, what is the probability that it lies in the span of some other vectors? In fact, once we write down a reasonable way of choosing a random real vector, if I choose a real vector in $\mathbb{R}^n$ at random, the probability that it lies in the span of $n-1$ other real vectors is 0!

23 Diagonalizing Symmetric Matrices

Symmetric matrices have many applications. For example, if we consider the shortest distance between pairs of important cities, we might get a table like this:

	Davis	Seattle	San Francisco
Davis	0	2000	80
Seattle	2000	0	2010
San Francisco	80	2010	0

Encoded as a matrix, we obtain:

$$M = \begin{pmatrix} 0 & 2000 & 80 \\ 2000 & 0 & 2010 \\ 80 & 2010 & 0 \end{pmatrix} = M^T.$$

Definition A matrix is *symmetric* if it obeys

$$M = M^T.$$

One very nice property of symmetric matrices is that they always have real eigenvalues. The general proof is an exercise, but here's an example for 2×2 matrices.

Example For a general symmetric 2×2 matrix, we have:

$$\begin{aligned} P_\lambda \begin{pmatrix} a & b \\ b & d \end{pmatrix} &= \det \begin{pmatrix} \lambda - a & -b \\ -b & \lambda - d \end{pmatrix} \\ &= (\lambda - a)(\lambda - d) - b^2 \\ &= \lambda^2 - (a+d)\lambda - b^2 + ad \\ \Rightarrow \lambda &= \frac{a+d}{2} \pm \sqrt{b^2 + \left(\frac{a-d}{2}\right)^2}. \end{aligned}$$

Notice that the discriminant $4b^2 + (a-d)^2$ is always positive, so that the eigenvalues must be real.

Now, suppose a symmetric matrix M has two distinct eigenvalues $\lambda \neq \mu$ and eigenvectors x and y:

$$Mx = \lambda x, \qquad My = \mu y.$$

Consider the dot product $x \cdot y = x^T y = y^T x$. And now calculate:

$$\begin{aligned}
x^T M y &= x^T \mu y = \mu x \cdot y, \text{ and} \\
x^T M y &= (y^T M x)^T \text{ (by transposing a } 1 \times 1 \text{ matrix)} \\
&= x^T M^T y \\
&= x^T M y \\
&= x^T \lambda y \\
&= \lambda x \cdot y.
\end{aligned}$$

Subtracting these two results tells us that:

$$0 = x^T M y - x^T M y = (\mu - \lambda)\, x \cdot y.$$

Since μ and λ were assumed to be distinct eigenvalues, $\lambda - \mu$ is non-zero, and so $x \cdot y = 0$. Then we have proved the following theorem.

Theorem 23.1. *Eigenvectors of a symmetric matrix with distinct eigenvalues are orthogonal.*

Reading homework: problem 23.1

Example The matrix $M = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ has eigenvalues determined by

$$\det(M - \lambda) = (2 - \lambda)^2 - 1 = 0.$$

Then the eigenvalues of M are 3 and 1, and the associated eigenvectors turn out to be $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. It is easily seen that these eigenvectors are orthogonal:

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0$$

In Lecture 21 we saw that the matrix P built from orthonormal basis vectors $\{v_1, \ldots, v_n\}$

$$P = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$$

was an orthogonal matrix:

$$P^{-1} = P^T, \text{ or } PP^T = I = P^T P.$$

Moreover, given any (unit) vector x_1, one can always find vectors $x_2, \ldots, x_n$ such that $\{x_1, \ldots, x_n\}$ is an orthonormal basis. (Such a basis can be obtained using the Gram-Schmidt procedure.)

Now suppose M is a symmetric $n \times n$ matrix and λ_1 is an eigenvalue with eigenvector x_1. Let the square matrix of column vectors P be the following:

$$P = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix},$$

where x_1 through x_n are orthonormal, and x_1 is an eigenvector for M, but the others are not necessarily eigenvectors for M. Then

$$MP = \begin{pmatrix} \lambda_1 x_1 & Mx_2 & \cdots & Mx_n \end{pmatrix}.$$

But P is an orthogonal matrix, so $P^{-1} = P^T$. Then:

$$\begin{aligned}
P^{-1} = P^T &= \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} \\
\Rightarrow P^T M P &= \begin{pmatrix} x_1^T \lambda_1 x_1 & * & \cdots & * \\ x_2^T \lambda_1 x_1 & * & \cdots & * \\ \vdots & & & \vdots \\ x_n^T \lambda_1 x_1 & * & \cdots & * \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & * & & \vdots \\ 0 & * & \cdots & * \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \hat{M} & \\ 0 & & & \end{pmatrix}
\end{aligned}$$

The last equality follows since $P^T M P$ is symmetric. The asterisks in the matrix are where "stuff" happens; this extra information is denoted by $\hat{M}$ in the final equation. We know nothing about $\hat{M}$ except that it is an $(n-1) \times (n-1)$ matrix and that it is symmetric. But then, by finding an (unit) eigenvector for $\hat{M}$, we could repeat this procedure successively. The end result would be a diagonal matrix with eigenvalues of M on the diagonal. Then we have proved a theorem.

Theorem 23.2. *Every symmetric matrix is similar to a diagonal matrix of its eigenvalues. In other words,*

$$M = M^T \Rightarrow M = PDP^T$$

where P is an orthogonal matrix and D is a diagonal matrix whose entries are the eigenvalues of M.

Reading homework: problem 23.2

To diagonalize a real symmetric matrix, begin by building an orthogonal matrix from an orthonormal basis of eigenvectors.

Example The symmetric matrix $M = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ has eigenvalues 3 and 1 with eigenvectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ respectively. From these eigenvectors, we normalize and build the orthogonal matrix:

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$$

Notice that $P^T P = I_2$. Then:

$$MP = \begin{pmatrix} \frac{3}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}.$$

In short, $MP = DP$, so $D = P^T MP$. Then D is the diagonalized form of M and P the associated change-of-basis matrix from the standard basis to the basis of eigenvectors.

References

Hefferon, Chapter Three, Section V: Change of Basis
Beezer, Chapter E, Section PEE, Subsection EHM
Beezer, Chapter E, Section SD, Subsection D
Wikipedia:

- Symmetric Matrix
- Diagonalizable Matrix
- Similar Matrix

Review Questions

1. (On Reality of Eigenvectors)

   (a) Suppose $z = x + iy$ where $x, y \in \mathbb{R}, i = \sqrt{-1}$, and $\overline{z} = x - iy$. Compute $z\overline{z}$ and $\overline{z}z$ in terms of x and y. What kind of numbers are $z\overline{z}$ and $\overline{z}z$? (The complex number $\overline{z}$ is called the *complex conjugate* of z).

   (b) Suppose that $\lambda = x + iy$ is a complex number with $x, y \in \mathbb{R}$, and that $\lambda = \overline{\lambda}$. Does this determine the value of x or y? What kind of number must λ be?

   (c) Let $x = \begin{pmatrix} z^1 \\ \vdots \\ z^n \end{pmatrix} \in \mathbb{C}^n$. Let $x^\dagger = \begin{pmatrix} \overline{z^1} & \cdots & \overline{z^n} \end{pmatrix} \in \mathbb{C}^n$. Compute $x^\dagger x$. Using the result of part 1a, what can you say about the number $x^\dagger x$? (E.g., is it real, imaginary, positive, negative, etc.)

   (d) Suppose $M = M^T$ is an $n \times n$ symmetric matrix with real entries. Let λ be an eigenvalue of M with eigenvector x, so $Mx = \lambda x$. Compute:
   $$\frac{x^\dagger M x}{x^\dagger x}$$

   (e) Suppose Λ is a 1×1 matrix. What is Λ^T?

   (f) What is the size of the matrix $x^\dagger M x$?

   (g) For any matrix (or vector) N, we can compute $\overline{N}$ by applying complex conjugation to each entry of N. Compute $\overline{(x^\dagger)^T}$. Then compute $\overline{(x^\dagger M x)^T}$.

   (h) Show that $\lambda = \overline{\lambda}$. Using the result of a previous part of this problem, what does this say about λ?

2. Let $x_1 = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$, where $a^2 + b^2 + c^2 = 1$. Find vectors x_2 and x_3 such that $\{x_1, x_2, x_3\}$ is an orthonormal basis for $\mathbb{R}^3$.

3. (Dimensions of Eigenspaces)

   (a) Let $A = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & -2 \\ 0 & -2 & 2 \end{pmatrix}$. Find all eigenvalues of A.

   (b) Find a basis for each eigenspace of A. What is the sum of the dimensions of the eigenspaces of A?

   (c) Based on your answer to the previous part, guess a formula for the sum of the dimensions of the eigenspaces of a real $n \times n$ symmetric matrix. Explain why your formula must work for any real $n \times n$ symmetric matrix.

24 Kernel, Range, Nullity, Rank

Given a linear transformation $L \colon V \to W$, we would like to know whether it has an inverse. That is, we would like to know whether there exists a linear transformation $M \colon W \to V$ such that for any vector $v \in V$, we have $M(L(v)) = v$, and for any vector $w \in W$, we have $L(M(w)) = w$. A linear transformation is just a special kind of function from one vector space to another. So before we discuss which linear transformations have inverses, let us first discuss inverses of arbitrary functions. When we later specialize to linear transformations, we'll also find some nice ways of creating subspaces.

Let $f \colon S \to T$ be a function from a set S to a set T. Recall that S is called the *domain* of f, T is called the *codomain* of f, and the set

$$\operatorname{ran}(f) = \operatorname{im}(f) = f(S) = \{f(s) | s \in S\} \subset T$$

is called the *range* or *image* of f. The image of f is the set of elements of T to which the function f maps, i.e., the things in T which you can get to by starting in S and applying f. We can also talk about the pre-image of any subset $U \subset T$:

$$f^{-1}(U) = \{s \in S | f(s) \in U\} \subset S.$$

The pre-image of a set U is the set of all elements of S which map to U.

The function f is *one-to-one* if different elements in S always map to different elements in T. That is, f is one-to-one if for any elements $x \neq y \in S$, we have that $f(x) \neq f(y)$. One-to-one functions are also called *injective* functions. Notice that injectivity is a condition on the pre-image of f.

The function f is *onto* if every element of T is mapped to by some element of S. That is, f is onto if for any $t \in T$, there exists some $s \in S$ such that $f(s) = t$. Onto functions are also called *surjective* functions. Notice that surjectivity is a condition on the image of f.

If f is both injective and surjective, it is *bijective*.

Theorem 24.1. *A function $f \colon S \to T$ has an inverse function $g \colon T \to S$ if and only if it is bijective.*

Proof. Suppose that f is bijective. Since f is surjective, every element $t \in T$ has at least one pre-image, and since f is injective, every t has no more than one pre-image. Therefore, to construct an inverse function g, we simply define $g(t)$ to be the unique pre-image $f^{-1}(t)$ of t.

Conversely, suppose that f has an inverse function g.

- The function f is injective:

  Suppose that we have $x, y \in S$ such that $f(x) = f(y)$. We must have that $g(f(s)) = s$ for any $s \in S$, so in particular $g(f(x)) = x$ and $g(f(y)) = y$. But since $f(x) = f(y)$, we have $g(f(x)) = g(f(y))$ so $x = y$. Therefore, f is injective.

- The function f is surjective:

  Let t be any element of T. We must have that $f(g(t)) = t$. Thus, $g(t)$ is an element of S which maps to t. So f is surjective.

□

Now let us restrict to the case that our function f is not just an arbitrary function, but a linear transformation between two vector spaces. Everything we said above for arbitrary functions is exactly the same for linear transformations. However, the linear structure of vector spaces lets us say much more about one-to-one and onto functions than we can say about functions on general sets. For example, we always know that a linear function sends 0_V to 0_W. You will show that a linear transformation is one-to-one if and only if 0_V is the only vector that is sent to 0_W: by looking at just one (very special) vector, we can figure out whether f is one-to-one. For arbitrary functions between arbitrary sets, things aren't nearly so convenient!

Let $L \colon V \to W$ be a linear transformation. Suppose L is *not* injective. Then we can find $v_1 \neq v_2$ such that $Lv_1 = Lv_2$. Then $v_1 - v_2 \neq 0$, but

$$L(v_1 - v_2) = 0.$$

Definition Let $L \colon V \to W$ be a linear transformation. The set of all vectors v such that $Lv = 0_W$ is called the *kernel of* L:

$$\ker L = \{v \in V | Lv = 0_W\}.$$

Theorem 24.2. *A linear transformation L is injective if and only if*

$$\ker L = \{0_V\}\,.$$

Proof. The proof of this theorem is an exercise. □

Notice that if L has matrix M in some basis, then finding the kernel of L is equivalent to solving the homogeneous system

$$MX = 0.$$

Example Let $L(x,y) = (x+y, x+2y, y)$. Is L one-to-one?

To find out, we can solve the linear system:

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right).$$

Then all solutions of $MX = 0$ are of the form $x = y = 0$. In other words, $\ker L = 0$, and so L is injective.

Reading homework: problem 24.1

Theorem 24.3. *Let $L\colon V \to W$. Then $\ker L$ is a subspace of V.*

Proof. Notice that if $L(v) = 0$ and $L(u) = 0$, then for any constants c, d, $L(cu + dv) = 0$. Then by the subspace theorem, the kernel of L is a subspace of V. □

This theorem has an interpretation in terms of the eigenspaces of $L\colon V \to V$. Suppose L has a zero eigenvalue. Then the associated eigenspace consists of all vectors v such that $Lv = 0v = 0$; in other words, the 0-eigenspace of L is exactly the kernel of L.

Returning to the previous example, let $L(x,y) = (x+y, x+2y, y)$. L is clearly not surjective, since L sends $\mathbb{R}^2$ to a plane in $\mathbb{R}^3$.

Example Let $L\colon \mathbb{R}^3 \to \mathbb{R}$ be the linear transformation defined by $L(x,y,z) = (x + y + z)$. Then $\ker L$ consists of all vectors $(x,y,z) \in \mathbb{R}^3$ such that $x + y + z = 0$. Therefore, the set

$$V = \{(x,y,z) \in \mathbb{R}^3 \mid x + y + z = 0\}$$

is a subspace of $\mathbb{R}^3$.

Notice that if $x = L(v)$ and $y = L(u)$, then for any constants c, d, $cx + dy = L(cv + du)$. Now the subspace theorem strikes again, and we have the following theorem.

Theorem 24.4. *Let $L\colon V \to W$. Then the image $L(V)$ is a subspace of W.*

To find a basis of the image of L, we can start with a basis $S = \{v_1, \ldots, v_n\}$ for V, and conclude (see the Review Exercises) that

$$L(V) = \operatorname{span} L(S) = \operatorname{span}\{L(v_1), \ldots, L(v_n)\}.$$

However, the set $\{L(v_1), \ldots, L(v_n)\}$ may not be linearly independent, so we solve

$$c^1 L(v_1) + \cdots + c^n L(v_n) = 0.$$

By finding relations amongst $L(S)$, we can discard vectors until a basis is arrived at. The size of this basis is the dimension of the image of L, which is known as the *rank* of L.

Definition The *rank* of a linear transformation L is the dimension of its image, written $\operatorname{rank} L = \dim L(V) = \dim \ \operatorname{ran} \ L$.

The *nullity* of a linear transformation is the dimension of the kernel, written $\operatorname{null} L = \dim \ker L$.

Theorem 24.5 (Dimension Formula)**.** *Let $L\colon V \to W$ be a linear transformation, with V a finite-dimensional vector space*[14]*. Then:*

$$\begin{aligned} \dim V &= \dim \ker V + \dim L(V) \\ &= \operatorname{null} L + \operatorname{rank} L. \end{aligned}$$

Proof. Pick a basis for V:

$$\{v_1, \ldots, v_p, u_1, \ldots, u_q\},$$

where $v_1, \ldots, v_p$ is also a basis for $\ker L$. This can always be done, for example, by finding a basis for the kernel of L and then extending to a basis for V. Then $p = \operatorname{null} L$ and $p + q = \dim V$. Then we need to show that $q = \operatorname{rank} L$. To accomplish this, we show that $\{L(u_1), \ldots, L(u_q)\}$ is a basis for $L(V)$.

To see that $\{L(u_1), \ldots, L(u_q)\}$ spans $L(V)$, consider any vector w in $L(V)$. Then we can find constants c^i, d^j such that:

$$\begin{aligned} w &= L(c^1 v_1 + \cdots + c^p v_p + d^1 u_1 + \cdots + d^q u_q) \\ &= c^1 L(v_1) + \cdots + c^p L(v_p) + d^1 L(u_1) + \cdots + d^q L(u_q) \\ &= d^1 L(u_1) + \cdots + d^q L(u_q) \text{ since } L(v_i) = 0, \\ \Rightarrow L(V) &= \operatorname{span}\{L(u_1), \ldots, L(u_q)\}. \end{aligned}$$

[14] The formula still makes sense for infinite dimensional vector spaces, such as the space of all polynomials, but the notion of a basis for an infinite dimensional space is more sticky than in the finite-dimensional case. Furthermore, the dimension formula for infinite dimensional vector spaces isn't useful for computing the rank of a linear transformation, since an equation like $\infty = \infty + x$ cannot be solved for x. As such, the proof presented assumes a finite basis for V.

Now we show that $\{L(u_1), \ldots, L(u_q)\}$ is linearly independent. We argue by contradiction: Suppose there exist constants d^j (not all zero) such that

$$\begin{aligned} 0 &= d^1L(u_1) + \cdots + d^qL(u_q) \\ &= L(d^1u_1 + \cdots + d^qu_q). \end{aligned}$$

But since the u^j are linearly independent, then $d^1u_1 + \cdots + d^qu_q \neq 0$, and so $d^1u_1 + \cdots + d^qu_q$ is in the kernel of L. But then $d^1u_1 + \cdots + d^qu_q$ must be in the span of $\{v_1, \ldots, v_p\}$, since this was a basis for the kernel. This contradicts the assumption that $\{v_1, \ldots, v_p, u_1, \ldots, u_q\}$ was a basis for V, so we are done. □

Reading homework: problem 24.2

24.1 Summary

We have seen that a linear transformation has an inverse if and only if it is bijective (i.e., one-to-one and onto). We also know that linear transformations can be represented by matrices, and we have seen many ways to tell whether a matrix is invertible. Here is a list of them.

Theorem 24.6 (Invertibility)**.** *Let M be an $n \times n$ matrix, and let $L\colon \mathbb{R}^n \to \mathbb{R}^n$ be the linear transformation defined by $L(v) = Mv$. Then the following statements are equivalent:*

1. *If V is any vector in $\mathbb{R}^n$, then the system $MX = V$ has exactly one solution.*
2. *The matrix M is row-equivalent to the identity matrix.*
3. *If v is any vector in $\mathbb{R}^n$, then $L(x) = v$ has exactly one solution.*
4. *The matrix M is invertible.*
5. *The homogeneous system $MX = 0$ has no non-zero solutions.*
6. *The determinant of M is not equal to 0.*
7. *The transpose matrix M^T is invertible.*
8. *The matrix M does not have 0 as an eigenvalue.*
9. *The linear transformation L does not have 0 as an eigenvalue.*
10. *The characteristic polynomial $\det(\lambda I - M)$ does not have 0 as a root.*
11. *The columns (or rows) of M span $\mathbb{R}^n$.*
12. *The columns (or rows) of M are linearly independent.*
13. *The columns (or rows) of M are a basis for $\mathbb{R}^n$.*
14. *The linear transformation L is injective.*
15. *The linear transformation L is surjective.*
16. *The linear transformation L is bijective.*

Note: it is important that M be an $n \times n$ matrix! If M is not square, then it can't be invertible, and many of the statements above are no longer equivalent to each other.

Proof. Many of these equivalences were proved earlier in these notes. Some were left as review questions or sample final questions. The rest are left as exercises for the reader. $\square$

References

Hefferon, Chapter Three, Section II.2: Rangespace and Nullspace (Recall that "homomorphism" is is used instead of "linear transformation" in Hefferon.)
Beezer, Chapter LT, Sections ILT-IVLT
Wikipedia:

- Rank
- Dimension Theorem
- Kernel of a Linear Operator

Review Questions

1. Let $L\colon V \to W$ be a linear transformation. Show that $\ker L = \{0_V\}$ if and only if L is one-to-one:

   (a) First, suppose that $\ker L = \{0_V\}$. Show that L is one-to-one. Think about methods of proof–does a proof by contradiction, a proof by induction, or a direct proof seem most appropriate?

   (b) Now, suppose that L is one-to-one. Show that $\ker L = \{0_V\}$. That is, show that 0_V is in $\ker L$, and then show that there are no other vectors in $\ker L$.

2. Let $\{v_1, \ldots, v_n\}$ be a basis for V. Explain why

$$L(V) = \operatorname{span}\{L(v_1), \ldots, L(v_n)\}.$$

3. Suppose $L\colon \mathbb{R}^4 \to \mathbb{R}^3$ whose matrix M in the standard basis is row equivalent to the following matrix:

$$\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

   Explain why the first three columns of the original matrix M form a basis for $L(\mathbb{R}^4)$.

Find and describe and algorithm (*i.e.* a general procedure) for finding a basis for $L(\mathbb{R}^n)$ when $L\colon \mathbb{R}^n \to \mathbb{R}^m$.
Finally, use your algorithm to find a basis for $L(\mathbb{R}^4)$ when $L\colon \mathbb{R}^4 \to \mathbb{R}^3$ is the linear transformation whose matrix M in the standard basis is

$$\begin{pmatrix} 2 & 1 & 1 & 4 \\ 0 & 1 & 0 & 5 \\ 4 & 1 & 1 & 6 \end{pmatrix}.$$

4. Claim: If $\{v_1, \ldots, v_n\}$ is a basis for $\ker L$, where $L\colon V \to W$, then it is always possible to extend this set to a basis for V.

   Choose a simple yet non-trivial linear transformation with a non-trivial kernel and verify the above claim for the transformation you choose.

5. Let $P_n(x)$ be the space of polynomials in x of degree less than or equal to n, and consider the derivative operator $\frac{\partial}{\partial x}$. Find the dimension of the kernel and image of $\frac{\partial}{\partial x}$.

   Now, consider $P_2(x, y)$, the space of polynomials of degree two or less in x and y. (Recall that xy is degree two, y is degree one and x^2y is degree three, for example.) Let $L = \frac{\partial}{\partial x} + \frac{\partial}{\partial y}$. (For example, $L(xy) = \frac{\partial}{\partial x}(xy) + \frac{\partial}{\partial y}(xy) = y + x$.) Find a basis for the kernel of L. Verify the dimension formula in this case.

25 Least Squares

Consider the linear system $L(x) = v$, where $L\colon U \stackrel{\text{linear}}{\longrightarrow} W$, and $v \in W$ is given. As we have seen, this system may have no solutions, a unique solution, or a space of solutions. But if v is not in the range of L then there will *never* be any solutions for $L(x) = v$.

However, for many applications we do not need a exact solution of the system; instead, we try to find the best approximation possible. To do this, we try to find x that minimizes $||L(x) - v||$.

> "My work always tried to unite the Truth with the Beautiful, but when I had to choose one or the other, I usually chose the Beautiful."
>
> – Hermann Weyl.

This method has many applications, such as when trying to fit a (perhaps linear) function to a "noisy" set of observations. For example, suppose we measured the position of a bicycle on a racetrack once every five seconds. Our observations won't be exact, but so long as the observations are right on average, we can figure out a best-possible linear function of position of the bicycle in terms of time.

Suppose M is the matrix for L in some bases for U and W, and v and x are given by column vectors V and X in these bases. Then we need to approximate

$$MX - V \approx 0\,.$$

Note that if $\dim U = n$ and $\dim W = m$ then M can be represented by an $m \times n$ matrix and x and v as vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Thus, we can write $W = L(U) \oplus L(U)^\perp$. Then we can uniquely write $v = v^\parallel + v^\perp$, with $v^\parallel \in L(U)$ and $v^\perp \in L(U)^\perp$.

Then we should solve $L(u) = v^\parallel$. In components, $v^\perp$ is just $V - MX$, and is the part we will eventually wish to minimize.

In terms of M, recall that $L(V)$ is spanned by the columns of M. (In the natural basis, the columns of M are $Me_1, \ldots, Me_n$.) Then $v^\perp$ must be perpendicular to the columns of M. *i.e.*, $M^T(V - MX) = 0$, or

$$M^T MX = M^T V.$$

Solutions X to $M^T MX = M^T V$ are called *least squares* solutions to $MX = V$.

Notice that any solution X to $MX = V$ is a least squares solution. However, the converse is often false. In fact, the equation $MX = V$ may have no solutions at all, but still have least squares solutions to $M^T MX = M^T V$.

Observe that since M is an $m \times n$ matrix, then M^T is an $n \times m$ matrix. Then $M^T M$ is an $n \times n$ matrix, and is symmetric, since $(M^T M)^T = M^T M$. Then, for any vector X, we can evaluate $X^T M^T MX$ to obtain a number. This is a very nice number, though! It is just the length $|MX|^2 = (MX)^T (MX) = X^T M^T MX$.

Reading homework: problem 25.1

Now suppose that $\ker L = \{0\}$, so that the only solution to $MX = 0$ is $X = 0$. (This need not mean that M is invertible because M is an $n \times m$ matrix, so not necessarily square.) However the square matrix $M^T M$ *is* invertible. To see this, suppose there was a vector X such that $M^T MX = 0$. Then it would follow that $X^T M^T MX = |MX|^2 = 0$. In other words the vector MX would have zero length, so could only be the zero vector. But we are assuming that $\ker L = \{0\}$ so $MX = 0$ implies $X = 0$. Thus the kernel of $M^T M$ is $\{0\}$ so this matrix is invertible. So, in this case, the least squares solution (the X that solves $M^T MX = MV$) is unique, and is equal to

$$X = (M^T M)^{-1} M^T V.$$

In a nutshell, this is the least squares method.

- Compute $M^T M$ and $M^T V$.
- Solve $(M^T M)X = M^T V$ by Gaussian elimination.

Example Captain Conundrum falls off of the leaning tower of Pisa and makes three (rather shaky) measurements of his velocity at three different times.

t s	v m/s
1	11
2	19
3	31

Having taken some calculus[15], he believes that his data are best approximated by a straight line

$$v = at + b.$$

[15] In fact, he is a *Calculus Superhero*.

Then he should find a and b to best fit the data.

$$\begin{aligned} 11 &= a \cdot 1 + b \\ 19 &= a \cdot 2 + b \\ 31 &= a \cdot 3 + b. \end{aligned}$$

As a system of linear equations, this becomes:

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} 11 \\ 19 \\ 31 \end{pmatrix}.$$

There is likely no actual straight line solution, so instead solve $M^T M X = M^T V$.

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 11 \\ 19 \\ 31 \end{pmatrix}.$$

This simplifies to the system:

$$\left(\begin{array}{cc|c} 14 & 6 & 142 \\ 6 & 3 & 61 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 10 \\ 0 & 1 & \frac{1}{3} \end{array}\right).$$

Then the least-squares fit is the line

$$v = 10\ t + \frac{1}{3}.$$

Notice that this equation implies that Captain Conundrum accelerates towards Italian soil at 10 m/s^2 (which is an excellent approximation to reality) and that he started at a downward velocity of $\frac{1}{3}$ m/s (perhaps somebody gave him a shove...)!

Congratulations, you have reached the end of these notes! You can test your skills on the sample final exam.

References

Hefferon, Chapter Three, Section VI.2: Gram-Schmidt Orthogonalization
Beezer, Part A, Section CF, Subsection DF
Wikipedia:

- Linear Least Squares
- Least Squares

Review Questions

1. Let $L : U \to V$ be a linear transformation. Suppose $v \in L(U)$ and you have found a vector u_{ps} that obeys $L(u_{\mathrm{ps}}) = v$.

   Explain why you need to compute $\ker L$ to describe the solution space of the linear system $L(u) = v$.

2. Suppose that M is an $m \times n$ matrix with trivial kernel. Show that for any vectors u and v in $\mathbb{R}^m$:

   - $u^T M^T M v = v^T M^T M u$
   - $v^T M^T M v \geq 0$.
   - If $v^T M^T M v = 0$, then $v = 0$.

   (Hint: Think about the dot product in $\mathbb{R}^n$.)

A Sample Midterm I Problems and Solutions

1. Solve the following linear system. Write the solution set in vector form. Check your solution. Write one particular solution and one homogeneous solution, if they exist. What does the solution set look like geometrically?

$$\begin{array}{rcrcrl} x & + & 3y & & & = 4 \\ x & - & 2y & + & z & = 1 \\ 2x & + & y & + & z & = 5 \end{array}$$

2. Consider the system

$$\left\{ \begin{array}{rcrcrcrcr} x & & & - & z & + & 2w & = & -1 \\ x & + & y & + & z & - & w & = & 2 \\ & - & y & - & 2z & + & 3w & = & -3 \\ 5x & + & 2y & - & z & + & 4w & = & 1 \end{array} \right.$$

   (a) Write an augmented matrix for this system.

   (b) Use elementary row operations to find its reduced row echelon form.

   (c) Write the solution set for the system in the form

$$\{X = X_0 + \sum_i \mu_i Y_i : \mu_i \in \mathbb{R}\}.$$

   (d) What are the vectors X_0 and Y_i called *and* which matrix equations do they solve?

   (e) Check separately that X_0 and each Y_i solve the matrix systems you claimed they solved in part (d).

3. Use row operations to invert the matrix

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 7 & 11 \\ 3 & 7 & 14 & 25 \\ 4 & 11 & 25 & 50 \end{pmatrix}$$

4. Let $M = \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix}$. Calculate $M^T M^{-1}$. Is M symmetric? What is the trace of the transpose of $f(M)$, where $f(x) = x^2 - 1$?

5. In this problem M is the matrix
$$M = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$
and X is the vector
$$X = \begin{pmatrix} x \\ y \end{pmatrix}.$$
Calculate all possible dot products between the vectors X and MX. Compute the lengths of X and MX. What is the angle between the vectors MX and X. Draw a picture of these vectors in the plane. For what values of θ do you expect equality in the triangle and Cauchy–Schwartz inequalities?

6. Let M be the matrix
$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$
Find a formula for M^k for any positive integer power k. Try some simple examples like $k = 2, 3$ if confused.

7. *Determinants:* The determinant $\det M$ of a 2×2 matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is defined by
$$\det M = ad - bc\,.$$
   (a) For which values of $\det M$ does M have an inverse?
   (b) Write down all 2×2 bit matrices with determinant 1. (Remember bits are either 0 or 1 and $1 + 1 = 0$.)
   (c) Write down all 2×2 bit matrices with determinant 0.

(d) Use one of the above examples to show why the following statement is FALSE.

Square matrices with the same determinant are always row equivalent.

8. What does it mean for a function to be linear? Check that integration is a linear function from V to V, where $V = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is integrable}\}$ is a vector space over $\mathbb{R}$ with usual addition and scalar multiplication.

9. What are the four main things we need to define for a vector space? Which of the following is a vector space over $\mathbb{R}$? For those that are not vector spaces, modify one part of the definition to make it into a vector space.

(a) $V = \{$ 2×2 matrices with entries in $\mathbb{R}\}$, usual matrix addition, and $k \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ka & b \\ kc & d \end{pmatrix}$ for $k \in \mathbb{R}$.

(b) $V = \{$polynomials with complex coefficients of degree $\leq 3\}$, with usual addition and scalar multiplication of polynomials.

(c) $V = \{$vectors in $\mathbb{R}^3$ with at least one entry containing a $1\}$, with usual addition and scalar multiplication.

10. *Subspaces:* If V is a vector space, we say that U is a *subspace* of V when the set U is also a vector space, using the vector addition and scalar multiplication rules of the vector space V. (Remember that $U \subset V$ says that "U is a subset of V", *i.e.*, all elements of U are also elements of V. The symbol $\forall$ means "for all" and $\in$ means "is an element of".)

Explain why additive closure ($u+w \in U \ \forall\ u, v \in U$) and multiplicative closure ($r.u \in U \ \forall\ r \in \mathbb{R},\ u \in V$) ensure that (i) the zero vector $0 \in U$ and (ii) every $u \in U$ has an additive inverse.

In fact it suffices to check closure under addition and scalar multiplication to verify that U is a vector space. Check whether the following choices of U are vector spaces:

(a) $U = \left\{ \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} : x, y \in \mathbb{R} \right\}$

(b) $U = \left\{ \begin{pmatrix} 1 \\ 0 \\ z \end{pmatrix} : z \in \mathbb{R} \right\}$

Solutions

1. *As an additional exercise, write out the row operations above the $\sim$ signs below:*

$$\left(\begin{array}{ccc|c} 1 & 3 & 0 & 4 \\ 1 & -2 & 1 & 1 \\ 2 & 1 & 1 & 5 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 3 & 0 & 4 \\ 0 & -5 & 1 & -3 \\ 0 & -5 & 1 & -3 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & \frac{3}{5} & \frac{11}{5} \\ 0 & 1 & -\frac{1}{5} & \frac{3}{5} \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Solution set

$$\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{11}{5} \\ \frac{3}{5} \\ 0 \end{pmatrix} + \mu \begin{pmatrix} -\frac{3}{5} \\ \frac{1}{5} \\ 1 \end{pmatrix} : \mu \in \mathbb{R} \right\}$$

Geometrically this represents a line in $\mathbb{R}^3$ through the point $\begin{pmatrix} \frac{11}{5} \\ \frac{3}{5} \\ 0 \end{pmatrix}$ and running parallel to the vector $\begin{pmatrix} -\frac{3}{5} \\ \frac{1}{5} \\ 1 \end{pmatrix}$.

A particular solution is $\begin{pmatrix} \frac{11}{5} \\ \frac{3}{5} \\ 0 \end{pmatrix}$ and a homogeneous solution is $\begin{pmatrix} -\frac{3}{5} \\ \frac{1}{5} \\ 1 \end{pmatrix}$.

As a double check note that

$$\begin{pmatrix} 1 & 3 & 0 \\ 1 & -2 & 1 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{11}{5} \\ \frac{3}{5} \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 5 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 3 & 0 \\ 1 & -2 & 1 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} -\frac{3}{5} \\ \frac{1}{5} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

2. (a) *Again, write out the row operations as an additional exercise.*

$$\left(\begin{array}{cccc|c} 1 & 0 & -1 & 2 & -1 \\ 1 & 1 & 1 & -1 & 2 \\ 0 & -1 & -2 & 3 & -3 \\ 5 & 2 & -1 & 4 & 1 \end{array}\right)$$

(b)

$$\sim \left(\begin{array}{rrrr|r} 1 & 0 & -1 & 2 & -1 \\ 0 & 1 & 2 & -3 & 3 \\ 0 & -1 & -2 & 3 & -3 \\ 0 & 2 & 4 & -6 & 6 \end{array}\right) \sim \left(\begin{array}{rrrr|r} 1 & 0 & -1 & 2 & -1 \\ 0 & 1 & 2 & -3 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

(c) Solution set

$$\left\{ X = \begin{pmatrix} -1 \\ 3 \\ 0 \\ 0 \end{pmatrix} + \mu_1 \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} + \mu_2 \begin{pmatrix} -2 \\ 3 \\ 0 \\ 1 \end{pmatrix} : \mu_1, \mu_2 \in \mathbb{R} \right\}.$$

(d) The vector $X_0 = \begin{pmatrix} -1 \\ 3 \\ 0 \\ 0 \end{pmatrix}$ is *a* particular solution and the vectors $Y_1 = \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix}$ and $Y_2 = \begin{pmatrix} -2 \\ 3 \\ 0 \\ 1 \end{pmatrix}$ are homogeneous solutions. Calling $M = \begin{pmatrix} 1 & 0 & -1 & 2 \\ 1 & 1 & 1 & -1 \\ 0 & -1 & -2 & 3 \\ 5 & 2 & -1 & 4 \end{pmatrix}$ and $V = \begin{pmatrix} -1 \\ 2 \\ -3 \\ 1 \end{pmatrix}$, they obey

$$MX = V\,, \qquad MY_1 = 0 = MY_2\,.$$

(e) This amounts to performing explicitly the matrix manipulations $MX - V$, MY_1, MY_2 and checking they all return the zero vector.

3. *As usual, be sure to write out the row operations above the* $\sim$*'s so your work can be easily checked.*

$$\left(\begin{array}{rrrr|rrrr} 1 & 2 & 3 & 4 & 1 & 0 & 0 & 0 \\ 2 & 4 & 7 & 11 & 0 & 1 & 0 & 0 \\ 3 & 7 & 14 & 25 & 0 & 0 & 1 & 0 \\ 4 & 11 & 25 & 50 & 0 & 0 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|cccc} 1 & 2 & 3 & 4 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 & -2 & 1 & 0 & 0 \\ 0 & 1 & 5 & 13 & -3 & 0 & 1 & 0 \\ 0 & 3 & 13 & 34 & -4 & 0 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|cccc} 1 & 0 & -7 & -22 & 7 & 0 & -2 & 0 \\ 0 & 1 & 5 & 13 & -3 & 0 & 1 & 0 \\ 0 & 0 & 1 & 3 & -2 & 1 & 0 & 0 \\ 0 & 0 & -2 & -5 & 5 & 0 & -3 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & -1 & -7 & 7 & -2 & 0 \\ 0 & 1 & 0 & -2 & 7 & -5 & 1 & 0 \\ 0 & 0 & 1 & 3 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 2 & -3 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & -6 & 9 & -5 & 1 \\ 0 & 1 & 0 & 0 & 9 & -1 & -5 & 2 \\ 0 & 0 & 1 & 0 & -5 & -5 & 9 & -3 \\ 0 & 0 & 0 & 1 & 1 & 2 & -3 & 1 \end{array}\right).$$

Check

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 7 & 11 \\ 3 & 7 & 14 & 25 \\ 4 & 11 & 25 & 50 \end{pmatrix} \begin{pmatrix} -6 & 9 & -5 & 1 \\ 9 & -1 & -5 & 2 \\ -5 & -5 & 9 & -3 \\ 1 & 2 & -3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

4.

$$M^T M^{-1} = \begin{pmatrix} 2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{3}{5} & -\frac{2}{5} \end{pmatrix} = \begin{pmatrix} \frac{11}{5} & -\frac{4}{5} \\ -\frac{2}{5} & \frac{3}{5} \end{pmatrix}.$$

Since $M^T M^{-1} \neq I$, it follows $M^T \neq M$ so M is *not* symmetric. Finally

$$\operatorname{tr} f(M)^T = \operatorname{tr} f(M) = \operatorname{tr}(M^2 - I) = \operatorname{tr} \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix} - \operatorname{tr} I$$

$$= (2 \cdot 2 + 1 \cdot 3) + (3 \cdot 1 + (-1) \cdot (-1)) - 2 = 9\,.$$

5. First

$$X \cdot (MX) = X^T M X = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

$$= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} x\cos\theta + y\sin\theta \\ -x\sin\theta + y\cos\theta \end{pmatrix} = (x^2+y^2)\cos\theta\,.$$

Now $||X|| = \sqrt{X \cdot X} = \sqrt{x^2+y^2}$ and $(MX)\cdot(MX) = XM^TMX$. But

$$M^TM = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$
$$= \begin{pmatrix} \cos^2\theta + \sin^2\theta & 0 \\ 0 & \cos^2\theta + \sin^2\theta \end{pmatrix} = I\,.$$

Hence $||MX|| = ||X|| = \sqrt{x^2+y^2}$. Thus the cosine of the angle between X and MX is given by

$$\frac{X \cdot (MX)}{||X||\;||MX||} = \frac{(x^2+y^2)\cos\theta}{\sqrt{x^2+y^2}\,\sqrt{x^2+y^2}} = \cos\theta\,.$$

In other words, the angle is θ OR $-\theta$. You should draw two pictures, one where the angle between X and MX is θ, the other where it is $-\theta$.

For Cauchy–Schwartz, $\frac{|X\cdot(MX)|}{||X||\;||MX||} = |\cos\theta| = 1$ when $\theta = 0, \pi$. For the triangle equality $MX = X$ achieves $||X + MX|| = ||X|| + ||MX||$, which requires $\theta = 0$.

6. This is a block matrix problem. Notice the that matrix M is really just $M = \begin{pmatrix} I & I \\ 0 & I \end{pmatrix}$, where I and 0 are the 3×3 identity zero matrices, respectively. But

$$M^2 = \begin{pmatrix} I & I \\ 0 & I \end{pmatrix} \begin{pmatrix} I & I \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & 2I \\ 0 & I \end{pmatrix}$$

and

$$M^3 = \begin{pmatrix} I & I \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 2I \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & 3I \\ 0 & I \end{pmatrix}$$

so, $M^k = \begin{pmatrix} I & kI \\ 0 & I \end{pmatrix}$, or explicitly

$$M^k = \begin{pmatrix} 1 & 0 & 0 & k & 0 & 0 \\ 0 & 1 & 0 & 0 & k & 0 \\ 0 & 0 & 1 & 0 & 0 & k \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

7. (a) Whenever $\det M = ad - bc \neq 0$.

   (b) Unit determinant bit matrices:

$$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}, \begin{pmatrix}1 & 1\\0 & 1\end{pmatrix}, \begin{pmatrix}1 & 0\\1 & 1\end{pmatrix}, \begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}, \begin{pmatrix}1 & 1\\1 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\1 & 1\end{pmatrix}.$$

   (c) Bit matrices with vanishing determinant:

$$\begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\0 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\1 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\0 & 1\end{pmatrix},$$

$$\begin{pmatrix}1 & 1\\0 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\1 & 1\end{pmatrix}, \begin{pmatrix}1 & 0\\1 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\0 & 1\end{pmatrix}, \begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}.$$

   As a check, count that the total number of 2×2 *bit matrices is* $2^{(\text{number of entries})} = 2^4 = 16$.

   (d) To disprove this statement, we just need to find a single counterexample. All the unit determinant examples above are actually row equivalent to the identity matrix, so focus on the bit matrices with vanishing determinant. Then notice (for example), that

$$\begin{pmatrix}1 & 1\\0 & 0\end{pmatrix} \not\sim \begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}.$$

   So we have found a pair of matrices that are not row equivalent but do have the same determinant. It follows that the statement is false.

8. We can call a function $f\colon V \longrightarrow W$ *linear* if the sets V and W are vector spaces and f obeys

$$f(\alpha u + \beta v) = \alpha f(u) + \beta f(v),$$

   for all $u, v \in V$ and $\alpha, \beta \in \mathbb{R}$.

   Now, integration is a linear transformation from the space V of all integrable functions (don't be confused between the definition of a linear function above, and integrable functions $f(x)$ which here are the vectors in V) to the real numbers $\mathbb{R}$, because $\int_{-\infty}^{\infty}(\alpha f(x) + \beta g(x))dx = \alpha \int_{-\infty}^{\infty} f(x)dx + \beta \int_{-\infty}^{\infty} g(x)dx$.

9. The four main ingredients are (i) a set V of vectors, (ii) a number field K (usually $K = \mathbb{R}$), (iii) a rule for adding vectors (vector addition) and (iv) a way to multiply vectors by a number to produce a new vector (scalar multiplication). There are, of course, ten rules that these four ingredients must obey.

   (a) This is not a vector space. Notice that distributivity of scalar multiplication requires $2u = (1+1)u = u + u$ for any vector u but

   $$2 \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2a & b \\ 2c & d \end{pmatrix}$$

   which does *not* equal

   $$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2a & 2b \\ 2c & 2d \end{pmatrix} .$$

   This could be repaired by taking

   $$k \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ka & kb \\ kc & kd \end{pmatrix} .$$

   (b) This is a vector space. *Although, the question does not ask you to, it is a useful exercise to verify that all ten vector space rules are satisfied.*

   (c) This is not a vector space for many reasons. An easy one is that $(1, -1, 0)$ and $(-1, 1, 0)$ are both in the space, but their sum $(0, 0, 0)$ is not (*i.e.*, additive closure fails). The easiest way to repair this would be to drop the requirement that there be at least one entry equaling 1.

10. (i) Thanks to multiplicative closure, if $u \in U$, so is $(-1) \cdot u$. But $(-1) \cdot u + u = (-1) \cdot u + 1 \cdot u = (-1+1) \cdot u = 0.u = 0$ (at each step in this chain of equalities we have used the fact that V is a vector space and therefore can use its vector space rules). In particular, this means that the zero vector of V is in U and is its zero vector also. (ii) Also, in V, for each u there is an element $-u$ such that $u + (-u) = 0$. But by additive close, $(-u)$ must also be in U, thus every $u \in U$ has an additive inverse.

(a) This is a vector space. First we check additive closure: let $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$ and $\begin{pmatrix} z \\ w \\ 0 \end{pmatrix}$ be arbitrary vectors in U. But since $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} + \begin{pmatrix} z \\ w \\ 0 \end{pmatrix} = \begin{pmatrix} x+z \\ y+w \\ 0 \end{pmatrix}$, so is their sum (because vectors in U are those whose third component vanishes). Multiplicative closure is similar: for any $\alpha \in \mathbb{R}$, $\alpha \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} \alpha x \\ \alpha y \\ 0 \end{pmatrix}$, which also has no third component, so is in U.

(b) This is not a vector space for various reasons. A simple one is that $u = \begin{pmatrix} 1 \\ 0 \\ z \end{pmatrix}$ is in U but the vector $u + u = \begin{pmatrix} 2 \\ 0 \\ 2z \end{pmatrix}$ is not in U (it has a 2 in the first component, but vectors in U always have a 1 there).

B Sample Midterm II Problems and Solutions

1. Find an LU decomposition for the matrix

$$\begin{pmatrix} 1 & 1 & -1 & 2 \\ 1 & 3 & 2 & 2 \\ -1 & -3 & -4 & 6 \\ 0 & 4 & 7 & -2 \end{pmatrix}$$

   Use your result to solve the system

$$\left\{ \begin{array}{rcrcrcrl} x & + & y & - & z & + & 2w & = 7 \\ x & + & 3y & + & 2z & + & 2w & = 6 \\ -x & - & 3y & - & 4z & + & 6w & = 12 \\ & & 4y & + & 7z & - & 2w & = -7 \end{array} \right.$$

2. Let

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

   Compute $\det A$. Find all solutions to (i) $AX = 0$ and (ii) $AX = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ for the vector $X \in \mathbb{R}^3$. Find, but do not solve, the characteristic polynomial of A.

3. Let M be any 2×2 matrix. Show

$$\det M = -\frac{1}{2}\mathrm{tr}M^2 + \frac{1}{2}(\mathrm{tr}M)^2 .$$

4. *The permanent:* Let $M = (M^i_j)$ be an $n \times n$ matrix. An operation producing a single number from M similar to the determinant is the "permanent"

$$\mathrm{perm}\, M = \sum_\sigma M^1_{\sigma(1)} M^2_{\sigma(2)} \cdots M^n_{\sigma(n)} .$$

For example

$$\operatorname{perm}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad + bc\,.$$

Calculate

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

What do you think would happen to the permanent of an $n \times n$ matrix M if (include a *brief* explanation with each answer):

(a) You multiplied M by a number λ.

(b) You multiplied a row of M by a number λ.

(c) You took the transpose of M.

(d) You swapped two rows of M.

5. Let X be an $n \times 1$ matrix subject to

$$X^T X = (1)\,,$$

and define

$$H = I - 2XX^T\,,$$

(where I is the $n \times n$ identity matrix). Show

$$H = H^T = H^{-1}.$$

6. Suppose λ is an eigenvalue of the matrix M with associated eigenvector v. Is v an eigenvector of M^k (where k is any positive integer)? If so, what would the associated eigenvalue be?

Now suppose that the matrix N is *nilpotent*, *i.e.*

$$N^k = 0$$

for some integer $k \geq 2$. Show that 0 is the only eigenvalue of N.

7. Let $M = \begin{pmatrix} 3 & -5 \\ 1 & -3 \end{pmatrix}$. Compute M^{12}. (Hint: $2^{12} = 4096$.)

8. *The Cayley Hamilton Theorem*: Calculate the characteristic polynomial $P_M(\lambda)$ of the matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Now compute the matrix polynomial $P_M(M)$. What do you observe? Now suppose the $n \times n$ matrix A is "similar" to a diagonal matrix D, in other words

$$A = P^{-1}DP$$

for some invertible matrix P and D is a matrix with values $\lambda_1, \lambda_2, \ldots \lambda_n$ along its diagonal. Show that the two matrix polynomials $P_A(A)$ and $P_A(D)$ are similar (*i.e.* $P_A(A) = P^{-1}P_A(D)P$). Finally, compute $P_A(D)$, what can you say about $P_A(A)$?

9. *Define* what it means for a set U to be a subspace of a vector space V. Now let U and W be subspaces of V. Are the following also subspaces? (Remember that $\cup$ means "union" and $\cap$ means "intersection".)

   (a) $U \cup W$

   (b) $U \cap W$

   In each case *draw* examples in $\mathbb{R}^3$ that justify your answers. If you answered "yes" to either part also give a general explanation why this is the case.

10. *Define* what it means for a set of vectors $\{v_1, v_2, \ldots, v_n\}$ to (i) be linearly independent, (ii) span a vector space V and (iii) be a basis for a vector space V.

    Consider the following vectors in $\mathbb{R}^3$

$$u = \begin{pmatrix} -1 \\ -4 \\ 3 \end{pmatrix}, \qquad v = \begin{pmatrix} 4 \\ 5 \\ 0 \end{pmatrix}, \qquad w = \begin{pmatrix} 10 \\ 7 \\ h+3 \end{pmatrix}.$$

    For which values of h is $\{u, v, w\}$ a basis for $\mathbb{R}^3$?

Solutions

1.

$$\begin{pmatrix} 1 & 1 & -1 & 2 \\ 1 & 3 & 2 & 2 \\ -1 & -3 & -4 & 6 \\ 0 & 4 & 7 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 2 & 3 & 0 \\ 0 & -2 & -5 & 8 \\ 0 & 4 & 7 & -2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & -1 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 2 & 3 & 0 \\ 0 & 0 & -2 & 8 \\ 0 & 0 & 1 & -2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & -1 & 1 & 0 \\ 0 & 2 & -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 2 & 3 & 0 \\ 0 & 0 & -2 & 8 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

To solve $MX = V$ using $M = LU$ we first solve $LW = V$ whose augmented matrix reads

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 7 \\ 1 & 1 & 0 & 0 & 6 \\ -1 & -1 & 1 & 0 & 12 \\ 0 & 2 & -\frac{1}{2} & 1 & -7 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 7 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 18 \\ 0 & 2 & -\frac{1}{2} & 1 & -7 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 7 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 18 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right),$$

from which we can read off W. Now we compute X by solving $UX = W$ with the augmented matrix

$$\left(\begin{array}{cccc|c} 1 & 1 & -1 & 2 & 7 \\ 0 & 2 & 3 & 0 & -1 \\ 0 & 0 & -2 & 8 & 18 \\ 0 & 0 & 0 & 2 & 4 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 1 & -1 & 2 & 7 \\ 0 & 2 & 3 & 0 & -1 \\ 0 & 0 & -2 & 0 & 2 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|c} 1 & 1 & -1 & 2 & 7 \\ 0 & 2 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right)$$

So $x = 1$, $y = 1$, $z = -1$ and $w = 2$.

2.

$$\det A = 1.(2.6 - 3.5) - 1.(2.6 - 3.4) + 1.(2.5 - 2.4) = -1\,.$$

(i) Since $\det A \neq 0$, the homogeneous system $AX = 0$ only has the solution $X = 0$. (ii) It is efficient to compute the adjoint

$$\text{adj } A = \begin{pmatrix} -3 & 0 & 2 \\ -1 & 2 & -1 \\ 1 & -1 & 0 \end{pmatrix}^T = \begin{pmatrix} -3 & -1 & 1 \\ 0 & 2 & -1 \\ 2 & -1 & 0 \end{pmatrix}$$

Hence

$$A^{-1} = \begin{pmatrix} 3 & 1 & -1 \\ 0 & -2 & 1 \\ -2 & 1 & 0 \end{pmatrix} .$$

Thus

$$X = \begin{pmatrix} 3 & 1 & -1 \\ 0 & -2 & 1 \\ -2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} .$$

Finally,

$$P_A(\lambda) = -\det \begin{pmatrix} 1-\lambda & 1 & 1 \\ 2 & 2-\lambda & 3 \\ 4 & 5 & 6-\lambda \end{pmatrix}$$

$$= -\Big[(1-\lambda)[(2-\lambda)(6-\lambda) - 15] - [2.(6-\lambda) - 12] + [10 - 4.(2-\lambda)]\Big]$$

$$= \lambda^3 - 9\lambda^2 - \lambda + 1 .$$

3. Call $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then $\det M = ad - bc$, yet

$$-\frac{1}{2}\operatorname{tr} M^2 + \frac{1}{2}(\operatorname{tr} M)^2 = -\frac{1}{2}\operatorname{tr}\begin{pmatrix} a^2 + bc & * \\ * & bc + d^2 \end{pmatrix} - \frac{1}{2}(a+d)^2$$

$$= -\frac{1}{2}(a^2 + 2bc + d^2) + \frac{1}{2}(a^2 + 2ad + d^2) = ad - bc ,$$

which is what we were asked to show.

4.

$$\text{perm} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = 1.(5.9 + 6.8) + 2.(4.9 + 6.7) + 3.(4.8 + 5.7) = 450 .$$

(a) Multiplying M by λ replaces every matrix element $M^i_{\sigma(j)}$ in the formula for the permanent by $\lambda M^i_{\sigma(j)}$, and therefore produces an overall factor λ^n.

(b) Multiplying the i^{th} row by λ replaces $M^i_{\sigma(j)}$ in the formula for the permanent by $\lambda M^i_{\sigma(j)}$. Therefore the permanent is multiplied by an overall factor λ.

(c) The permanent of a matrix transposed equals the permanent of the original matrix, because in the formula for the permanent this amounts to summing over permutations of rows rather than columns. But we could then sort the product $M_1^{\sigma(1)} M_2^{\sigma(2)} \ldots M_n^{\sigma(n)}$ back into its original order using the inverse permutation σ^{-1}. But summing over permutations is equivalent to summing over inverse permutations, and therefore the permanent is unchanged.

(d) Swapping two rows also leaves the permanent unchanged. The argument is almost the same as in the previous part, except that we need only reshuffle two matrix elements $M^j_{\sigma(i)}$ and $M^i_{\sigma(j)}$ (in the case where rows i and j were swapped). Then we use the fact that summing over all permutations σ or over all permutations $\widetilde{\sigma}$ obtained by swapping a pair in σ are equivalent operations.

5. Firstly, lets call $(1) = 1$ (the 1×1 identity matrix). Then we calculate

$$H^T = (I - 2XX^T)^T = I^T - 2(XX^T)^T = I - 2(X^T)^T X^T = I - 2XX^T = H\,,$$

which demonstrates the first equality. Now we compute

$$H^2 = (I - 2XX^T)(I - 2XX^T) = I - 4XX^T + 4XX^TXX^T$$

$$= I - 4XX^T + 4X(X^TX)X^T = I - 4XX^T + 4X.1.X^T = I\,.$$

So, since $HH = I$, we have $H^{-1} = H$.

6. We know $Mv = \lambda v$. Hence

$$M^2 v = MMv = M\lambda v = \lambda M v = \lambda^2 v\,,$$

and similarly

$$M^k v = \lambda M^{k-1} v = \ldots = \lambda^k v\,.$$

So v is an eigenvector of M^k with eigenvalue λ^k.

Now let us assume v is an eigenvector of the nilpotent matrix N with eigenvalue λ. Then from above

$$N^k v = \lambda^k v$$

but by nilpotence, we also have

$$N^k v = 0$$

Hence $\lambda^k v = 0$ and v (being an eigenvector) cannot vanish. Thus $\lambda^k = 0$ and in turn $\lambda = 0$.

7. Let us think about the eigenvalue problem $Mv = \lambda v$. This has solutions when

$$0 = \det \begin{pmatrix} 3-\lambda & -5 \\ 1 & -3-\lambda \end{pmatrix} = \lambda^2 - 4 \Rightarrow \lambda = \pm 2\,.$$

The associated eigenvalues solve the homogeneous systems (in augmented matrix form)

$$\left(\begin{array}{cc|c} 1 & -5 & 0 \\ 1 & -5 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & -5 & 0 \\ 0 & 0 & 0 \end{array}\right) \text{ and } \left(\begin{array}{cc|c} 5 & -5 & 0 \\ 1 & -1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right),$$

respectively, so are $v_2 = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$ and $v_{-2} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Hence $M^{12} v_2 = 2^{12} v_2$ and $M^{12} v_{-2} = (-2)^{12} v_{-2}$. Now, $\begin{pmatrix} x \\ y \end{pmatrix} = \frac{x-y}{4} \begin{pmatrix} 5 \\ 1 \end{pmatrix} - \frac{x-5y}{4} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ (this was obtained by solving the linear system $av_2 + bv_{-2} =$ for a and b). Thus

$$M \begin{pmatrix} x \\ y \end{pmatrix} = \frac{x-y}{4} M v_2 - \frac{x-5y}{4} M v_{-2}$$

$$= 2^{12} \Big(\frac{x-y}{4} v_2 - \frac{x-5y}{4} v_{-2} \Big) = 2^{12} \begin{pmatrix} x \\ y \end{pmatrix}\,.$$

Thus

$$M^{12} = \begin{pmatrix} 4096 & 0 \\ 0 & 4096 \end{pmatrix}\,.$$

If you understand the above explanation, then you have a good understanding of diagonalization. A quicker route is simply to observe that $M^2 = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$.

8.
$$P_M(\lambda) = (-1)^2 \det \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} = (\lambda - a)(\lambda - d) - bc\,.$$

Thus
$$P_M(M) = (M - aI)(M - dI) - bcI$$
$$= \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} - \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \right) \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} - \begin{pmatrix} d & 0 \\ 0 & d \end{pmatrix} \right) - \begin{pmatrix} bc & 0 \\ 0 & bc \end{pmatrix}$$
$$= \begin{pmatrix} 0 & b \\ c & d-a \end{pmatrix} \begin{pmatrix} a-d & b \\ c & 0 \end{pmatrix} - \begin{pmatrix} bc & 0 \\ 0 & bc \end{pmatrix} = 0\,.$$

Observe that any 2×2 matrix is a zero of its own characteristic polynomial (*in fact this holds for square matrices of any size*).

Now if $A = P^{-1}DP$ then $A^2 = P^{-1}DPP^{-1}DP = P^{-1}D^2P$. Similarly $A^k = P^{-1}D^kP$. So for *any* matrix polynomial we have

$$\begin{array}{rl} & A^n + c_1 A^{n-1} + \cdots c_{n-1}A + c_n I \\ = & P^{-1}D^nP + c_1 P^{-1}D^{n-1}P + \cdots c_{n-1}P^{-1}DP + c_n P^{-1}P \\ = & P^{-1}(D^n + c_1 D^{n-1} + \cdots c_{n-1}D + c_n I)P\,. \end{array}$$

Thus we may conclude $P_A(A) = P^{-1}P_A(D)P$.

Now suppose $D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \lambda_n \end{pmatrix}$. Then

$$P_A(\lambda) = \det(\lambda I - A) = \det(\lambda P^{-1}IP - P^{-1}DP) = \det P . \det(\lambda I - D) . \det P$$

$$= \det(\lambda I - D) = \det \begin{pmatrix} \lambda - \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda - \lambda_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \lambda - \lambda_n \end{pmatrix}$$
$$= (\lambda - \lambda_1)(\lambda - \lambda_2) \ldots (\lambda - \lambda_n)\,.$$

Thus we see that λ_1, $\lambda_2, \ldots, \lambda_n$ are the eigenvalues of M. Finally we compute
$$P_A(D) = (D - \lambda_1)(D - \lambda_2) \ldots (D - \lambda_n)$$

$$= \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \lambda_n \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & 0 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \lambda_n \end{pmatrix} \cdots \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & 0 \end{pmatrix} = 0\,.$$

We conclude the $P_M(M) = 0$.

9. A subset of a vector space is called a subspace if it itself is a vector space, using the rules for vector addition and scalar multiplication inherited from the original vector space.

   (a) So long as $U \neq U \cup W \neq W$ the answer is *no*. Take, for example, U to be the x-axis in $\mathbb{R}^2$ and W to be the y-axis. Then $\begin{pmatrix}1 & 0\end{pmatrix} \in U$ and $\begin{pmatrix}0 & 1\end{pmatrix} \in W$, but $\begin{pmatrix}1 & 0\end{pmatrix} + \begin{pmatrix}0 & 1 & 0\end{pmatrix} = \begin{pmatrix}1 & 1\end{pmatrix} \notin U \cup W$. So $U \cup W$ is not additively closed and is not a vector space (and thus not a subspace). It is easy to draw the example described.

   (b) Here the answer is always *yes*. The proof is not difficult. Take a vector u and w such that $u \in U \cap W \ni w$. This means that *both* u and w are in *both* U and W. But, since U is a vector space, $\alpha u + \beta w$ is also in U. Similarly, $\alpha u + \beta w \in W$. Hence $\alpha u + \beta w \in U \cap W$. So closure holds in $U \cap W$ and this set is a subspace by the subspace theorem. Here, a good picture to draw is two planes through the origin in $\mathbb{R}^3$ intersecting at a line (also through the origin).

10. (i) We say that the vectors $\{v_1, v_2, \ldots v_n\}$ are linearly independent if there exist *no* constants c^1, $c^2, \ldots c^n$ (all non-vanishing) such that $c^1 v_1 + c^2 v_2 + \cdots + c^n v_n = 0$. Alternatively, we can require that there is no non-trivial solution for scalars c^1, $c^2, \ldots, c^n$ to the linear system $c^1 v_1 + c^2 v_2 + \cdots + c^n v_n = 0$. (ii) We say that these vectors span a vector space V if the set $\operatorname{span}\{v_1, v_2, \ldots v_n\} = \{c^1 v_1 + c^2 v_2 + \cdots + c^n v_n : c^1, c^2, \ldots c^n \in \mathbb{R}\} = V$. (iii) We call $\{v_1, v_2, \ldots v_n\}$ a basis for V if $\{v_1, v_2, \ldots v_n\}$ are linearly independent *and* $\operatorname{span}\{v_1, v_2, \ldots v_n\} = V$.

    For u, v, w to be a basis for $\mathbb{R}^3$, we firstly need (the spanning requirement) that any vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ can be written as a linear combination of

u, v and w

$$c^1 \begin{pmatrix} -1 \\ -4 \\ 3 \end{pmatrix} + c^2 \begin{pmatrix} 4 \\ 5 \\ 0 \end{pmatrix} + c^3 \begin{pmatrix} 10 \\ 7 \\ h+3 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

The linear independence requirement implies that when $x = y = z = 0$, the only solution to the above system is $c^1 = c^2 = c^3 = 0$. But the above system in matrix language reads

$$\begin{pmatrix} -1 & 4 & 10 \\ -4 & 5 & 7 \\ 3 & 0 & h+3 \end{pmatrix} \begin{pmatrix} c^1 \\ c^2 \\ c^3 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

Both requirements mean that the matrix on the left hand side must be invertible, so we examine its determinant

$$\det \begin{pmatrix} -1 & 4 & 10 \\ -4 & 5 & 7 \\ 3 & 0 & h+3 \end{pmatrix} = -4.(-4.(h+3) - 7.3) + 5.(-1.(h+3) - 10.3)$$

$$= 11(h-3) .$$

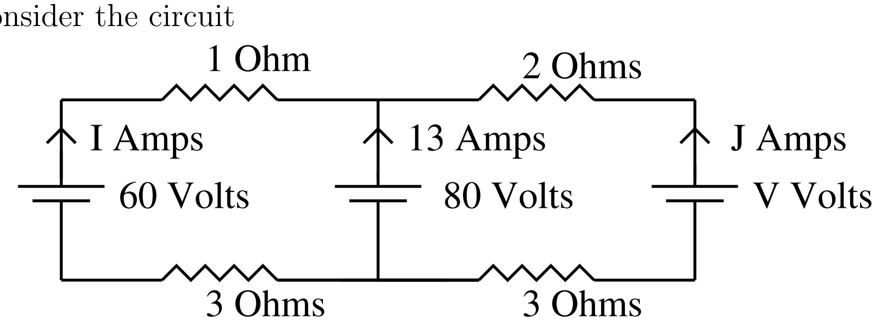
Hence we obtain a basis whenever $h \neq 3$.

C Sample Final Problems and Solutions

1. Define the following terms:

   (a) An *orthogonal matrix*.
   (b) A *basis* for a vector space.
   (c) The *span* of a set of vectors.
   (d) The *dimension* of a vector space.
   (e) An *eigenvector*.
   (f) A *subspace* of a vector space.
   (g) The *kernel* of a linear transformation.
   (h) The *nullity* of a linear transformation.
   (i) The *image* of a linear transformation.
   (j) The *rank* of a linear transformation.
   (k) The *characteristic polynomial* of a square matrix.
   (l) An *equivalence relation*.
   (m) A *homogeneous solution* to a linear system of equations.
   (n) A *particular solution* to a linear system of equations.
   (o) The *general solution* to a linear system of equations.
   (p) The *direct sum* of a pair of subspaces of a vector space.
   (q) The *orthogonal complement* to a subspace of a vector space.

2. *Kirchoff's laws*: Electrical circuits are easy to analyze using systems of equations. The change in voltage (measured in Volts) around any loop due to batteries $|\,|$ and resistors $/\!\backslash\!/\!\backslash\!/\!\backslash$ (given by the product of the current measured in Amps and resistance measured in Ohms) equals zero. Also, the sum of currents entering any junction vanishes. Consider the circuit



Find all possible equations for the unknowns I, J and V and then solve for I, J and V. Give your answers with correct units.

3. Suppose M is the matrix of a linear transformation

$$L : U \to V$$

and the vector spaces U and V have dimensions

$$\dim U = n\,, \qquad \dim V = m\,,$$

and

$$m \neq n\,.$$

Also assume

$$\ker L = \{0_U\}\,.$$

   (a) How many rows does M have?
   (b) How many columns does M have?
   (c) Are the columns of M linearly independent?
   (d) What size matrix is $M^T M$?
   (e) What size matrix is $M M^T$?
   (f) Is $M^T M$ invertible?
   (g) is $M^T M$ symmetric?
   (h) Is $M^T M$ diagonalizable?
   (i) Does $M^T M$ have a zero eigenvalue?
   (j) Suppose $U = V$ and $\ker L \neq \{0_U\}$. Find an eigenvalue of M.
   (k) Suppose $U = V$ and $\ker L \neq \{0_U\}$. Find $\det M$.

4. Consider the system of equations

$$\begin{array}{ccccccccc} x & + & y & + & z & + & w & = & 1 \\ x & + & 2y & + & 2z & + & 2w & = & 1 \\ x & + & 2y & + & 3z & + & 3w & = & 1 \end{array}$$

Express this system as a matrix equation $MX = V$ and then find the solution set by computing an LU decomposition for the matrix M (be sure to use back and forward substitution).

5. Compute the following determinants

$$\det\begin{pmatrix}1&2\\3&4\end{pmatrix},\ \det\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix},\ \det\begin{pmatrix}1&2&3&4\\5&6&7&8\\9&10&11&12\\13&14&15&16\end{pmatrix},$$

$$\det\begin{pmatrix}1&2&3&4&5\\6&7&8&9&10\\11&12&13&14&15\\16&17&18&19&20\\21&22&23&24&25\end{pmatrix}.$$

Now test your skills on

$$\det\begin{pmatrix}1&2&3&\cdots&n\\n+1&n+2&n+3&\cdots&2n\\2n+1&2n+2&2n+3&&3n\\\vdots&&&\ddots&\vdots\\n^2-n+1&n^2-1+2&n^2-n+3&\cdots&n^2\end{pmatrix}.$$

Make sure to jot down a few brief notes explaining any clever tricks you use.

6. For which values of a does

$$U=\operatorname{span}\left\{\begin{pmatrix}1\\0\\1\end{pmatrix},\begin{pmatrix}1\\2\\-3\end{pmatrix},\begin{pmatrix}a\\1\\0\end{pmatrix}\right\}=\mathbb{R}^3\ ?$$

For any special values of a at which $U\neq\mathbb{R}^3$, express the subspace U as the span of the least number of vectors possible. Give the dimension of U for these cases and draw a picture showing U inside $\mathbb{R}^3$.

7. *Vandermonde determinant:* Calculate the following determinants

$$\det\begin{pmatrix}1&x\\1&y\end{pmatrix},\quad \det\begin{pmatrix}1&x&x^2\\1&y&y^2\\1&z&z^2\end{pmatrix},\quad \det\begin{pmatrix}1&x&x^2&x^3\\1&y&y^2&y^3\\1&z&z^2&z^3\\1&w&w^2&w^3\end{pmatrix}.$$

Be sure to factorize you answers, if possible.

Challenging: Compute the determinant

$$\det\begin{pmatrix} 1 & x_1 & (x_1)^2 & \cdots & (x_1)^{n-1} \\ 1 & x_2 & (x_2)^2 & \cdots & (x_2)^{n-1} \\ 1 & x_3 & (x_3)^2 & \cdots & (x_3)^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & (x_n)^2 & \cdots & (x_n)^{n-1} \end{pmatrix}.$$

8. (a) Do the vectors $\left\{\begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}3\\2\\1\end{pmatrix}, \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$ form a basis for $\mathbb{R}^3$? *Be sure to justify your answer.*

   (b) Find a basis for $\mathbb{R}^4$ that includes the vectors $\begin{pmatrix}1\\2\\3\\4\end{pmatrix}$ and $\begin{pmatrix}4\\3\\2\\1\end{pmatrix}$.

   (c) Explain in words how to generalize your computation in part (b) to obtain a basis for $\mathbb{R}^n$ that includes a given pair of (linearly independent) vectors u and v.

9. Elite NASA engineers determine that if a satellite is placed in orbit starting at a point $\mathcal{O}$, it will return exactly to that same point after one orbit of the earth. Unfortunately, if there is a small mistake in the original location of the satellite, which the engineers label by a vector X in $\mathbb{R}^3$ with origin[16] at $\mathcal{O}$, after one orbit the satellite will instead return to some other point $Y \in \mathbb{R}^3$. The engineer's computations show that Y is related to X by a matrix

$$Y = \begin{pmatrix} 0 & \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} & 0 \end{pmatrix} X\,.$$

   (a) Find all eigenvalues of the above matrix.

[16] This is a spy satellite. The exact location of $\mathcal{O}$, the orientation of the coordinate axes in $\mathbb{R}^3$ and the unit system employed by the engineers are CIA secrets.

(b) Determine *all* possible eigenvectors associated with each eigenvalue.

Let us assume that the rule found by the engineers applies to all subsequent orbits. Discuss case by case, what will happen to the satellite if the initial mistake in its location is in a direction given by an eigenvector.

10. In this problem the scalars in the vector spaces are bits ($0, 1$ with $1+1=0$). The space B^k is the vector space of bit-valued, k-component column vectors.

(a) Find a basis for B^3.

(b) Your answer to part (a) should be a list of vectors v_1, $v_2, \ldots v_n$. What number did you find for n?

(c) How many elements are there in the *set* B^3.

(d) What is the dimension of the vector space B^3.

(e) Suppose $L : B^3 \to B = \{0, 1\}$ is a linear transformation. Explain why specifying $L(v_1)$, $L(v_2), \ldots, L(v_n)$ completely determines L.

(f) Use the notation of part (e) to list all linear transformations

$$L : B^3 \to B\,.$$

How many different linear transformations did you find? Compare your answer to part (c).

(g) Suppose $L_1 : B^3 \to B$ and $L_2 : B^3 \to B$ are linear transformations, and α and β are bits. Define a new map $(\alpha L_1 + \beta L_2) : B^3 \to B$ by

$$(\alpha L_1 + \beta L_2)(v) = \alpha L_1(v) + \beta L_2(v).$$

Is this map a linear transformation? Explain.

(h) Do you think the set of all linear transformations from B^3 to B is a vector space using the addition rule above? If you answer yes, give a basis for this vector space and state its dimension.

11. A team of distinguished, post-doctoral engineers analyzes the design for a bridge across the English channel. They notice that the force on the center of the bridge when it is displaced by an amount $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is given by

$$F = \begin{pmatrix} -x - y \\ -x - 2y - z \\ -y - z \end{pmatrix}.$$

Moreover, having read Newton's Principiæ, they know that force is proportional to acceleration so that[17]

$$F = \frac{d^2X}{dt^2}.$$

Since the engineers are worried the bridge might start swaying in the heavy channel winds, they search for an oscillatory solution to this equation of the form[18]

$$X = \cos(\omega t) \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

   (a) By plugging their proposed solution in the above equations the engineers find an eigenvalue problem

   $$M \begin{pmatrix} a \\ b \\ c \end{pmatrix} = -\omega^2 \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

   Here M is a 3×3 matrix. Which 3×3 matrix M did the engineers find? *Justify your answer.*

   (b) Find the eigenvalues and eigenvectors of the matrix M.

   (c) The number $|\omega|$ is often called a *characteristic frequency*. What characteristic frequencies do you find for the proposed bridge?

[17]The bridge is intended for French and English military vehicles, so the exact units, coordinate system and constant of proportionality are state secrets.

[18]Here, a, b, c and ω are constants which we aim to calculate.

(d) Find an orthogonal matrix P such that $MP = PD$ where D is a diagonal matrix. *Be sure to also state your result for D.*

(e) Is there a direction in which displacing the bridge yields no force? If so give a vector in that direction. *Briefly* evaluate the quality of this bridge design.

12. *Conic Sections*: The equation for the most general conic section is given by

$$ax^2 + 2bxy + dy^2 + 2cx + 2ey + f = 0\,.$$

Our aim is to analyze the solutions to this equation using matrices.

(a) Rewrite the above quadratic equation as one of the form

$$X^T MX + X^T C + C^T X + f = 0$$

relating an unknown column vector $X = \begin{pmatrix} x \\ y \end{pmatrix}$, its transpose X^T, a 2×2 matrix M, a constant column vector C and the constant f.

(b) Does your matrix M obey any special properties? Find its eigenvalues. You may call your answers λ and μ for the rest of the problem to save writing.

For the rest of this problem we will focus on central conics for which the matrix M is invertible.

(c) Your equation in part (a) above should be be quadratic in X. Recall that if $m \neq 0$, the quadratic equation $mx^2 + 2cx + f = 0$ can be rewritten by *completing the square*

$$m\Big(x + \frac{c}{m}\Big)^2 = \frac{c^2}{m} - f\,.$$

Being very careful that you are now dealing with matrices, use the same trick to rewrite your answer to part (a) in the form

$$Y^T MY = g.$$

Make sure you give formulas for the new unknown column vector Y and constant g in terms of X, M, C and f. You need not multiply out any of the matrix expressions you find.

If all has gone well, you have found a way to shift coordinates for the original conic equation to a new coordinate system with its origin at the center of symmetry. Our next aim is to rotate the coordinate axes to produce a readily recognizable equation.

(d) Why is the angle between vectors V and W is not changed when you replace them by PV and PW for P any orthogonal matrix?

(e) Explain how to choose an orthogonal matrix P such that $MP = PD$ where D is a diagonal matrix.

(f) For the choice of P above, define our final unknown vector Z by $Y = PZ$. Find an expression for $Y^T MY$ in terms of Z and the eigenvalues of M.

(g) Call $Z = \begin{pmatrix} z \\ w \end{pmatrix}$. What equation do z and w obey? (Hint, write your answer using λ, μ and g.)

(h) Central conics are circles, ellipses, hyperbolae or a pair of straight lines. Give examples of values of (λ, μ, g) which produce each of these cases.

13. Let $L\colon V \to W$ be a linear transformation between finite-dimensional vector spaces V and W, and let M be a matrix for L (with respect to some basis for V and some basis for W). We know that L has an inverse if and only if it is bijective, and we know a lot of ways to tell whether M has an inverse. In fact, L has an inverse if and only if M has an inverse:

    (a) Suppose that L is bijective (i.e., one-to-one and onto).

    i. Show that $\dim V = \operatorname{rank} L = \dim W$.
    ii. Show that 0 is not an eigenvalue of M.
    iii. Show that M is an invertible matrix.

    (b) Now, suppose that M is an invertible matrix.

    i. Show that 0 is not an eigenvalue of M.
    ii. Show that L is injective.
    iii. Show that L is surjective.

14. Captain Conundrum gives Queen Quandary a pair of newborn doves, male and female for her birthday. After one year, this pair of doves breed and produce a pair of dove eggs. One year later these eggs hatch yielding a new pair of doves while the original pair of doves breed again and an additional pair of eggs are laid. Captain Conundrum is very happy because now he will never need to buy the Queen a present ever again!

    Let us say that in year zero, the Queen has no doves. In year one she has one pair of doves, in year two she has two pairs of doves *etc...* Call F_n the number of pairs of doves in years n. For example, $F_0 = 0$, $F_1 = 1$ and $F_2 = 1$. Assume no doves die and that the same breeding pattern continues well into the future. Then $F_3 = 2$ because the eggs laid by the first pair of doves in year two hatch. Notice also that in year three, two pairs of eggs are laid (by the first and second pair of doves). Thus $F_4 = 3$.

    (a) Compute F_5 and F_6.

    (b) Explain why (for any $n \geq 2$) the following *recursion relation* holds
    $$F_n = F_{n-1} + F_{n-2}\,.$$

    (c) Let us introduce a column vector $X_n = \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}$. Compute X_1 and X_2. Verify that these vectors obey the relationship
    $$X_2 = MX_1 \text{ where } M = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\,.$$

    (d) Show that $X_{n+1} = MX_n$.

    (e) Diagonalize M. (*I.e.*, write M as a product $M = PDP^{-1}$ where D is diagonal.)

    (f) Find a simple expression for M^n in terms of P, D and P^{-1}.

    (g) Show that $X_{n+1} = M^n X_1$.

    (h) The number
    $$\varphi = \frac{1+\sqrt{5}}{2}$$
    is called the *golden ratio*. Write the eigenvalues of M in terms of φ.

(i) Put your results from parts (c), (f) and (g) together (along with a short matrix computation) to find the formula for the number of doves F_n in year n expressed in terms of φ, $1-\varphi$ and n.

15. Use Gram–Schmidt to find an orthonormal basis for

$$\operatorname{span}\left\{\begin{pmatrix}1\\1\\1\\1\end{pmatrix},\begin{pmatrix}1\\0\\1\\1\end{pmatrix},\begin{pmatrix}0\\0\\1\\2\end{pmatrix}\right\}.$$

16. Let M be the matrix of a linear transformation $L : V \to W$ in given bases for V and W. Fill in the blanks below with one of the following six vector spaces: V, W, $\ker L$, $\big(\ker L\big)^{\perp}$, $\operatorname{im} L$, $\big(\operatorname{im} L\big)^{\perp}$.

(a) The columns of M span ______ in the basis given for ______.

(b) The rows of M span ______ in the basis given for ______.

Suppose

$$M = \begin{pmatrix}1 & 2 & 1 & 3\\ 2 & 1 & -1 & 2\\ 1 & 0 & 0 & -1\\ 4 & 1 & -1 & 0\end{pmatrix}$$

is the matrix of L in the bases $\{v_1, v_2, v_3, v_4\}$ for V and $\{w_1, w_2, w_3, w_4\}$ for W. Find bases for $\ker L$ and $\operatorname{im} L$. Use the dimension formula to check your result.

17. Captain Conundrum collects the following data set

y	x
5	-2
2	-1
0	1
3	2

which he believes to be well-approximated by a parabola

$$y = ax^2 + bx + c\,.$$

(a) Write down a system of four linear equations for the unknown coefficients a, b and c.

(b) Write the augmented matrix for this system of equations.

(c) Find the reduced row echelon form for this augmented matrix.

(d) Are there any solutions to this system?

(e) Find the least squares solution to the system.

(f) What value does Captain Conundrum predict for y when $x = 2$?

18. Suppose you have collected the following data for an experiment

x	y
x_1	y_1
x_2	y_2
x_3	y_3

and believe that the result is well modeled by a straight line

$$y = mx + b\,.$$

(a) Write down a linear system of equations you could use to find the slope m and constant term b.

(b) Arrange the unknowns (m, b) in a column vector X and write your answer to (a) as a matrix equation

$$MX = V\,.$$

Be sure to give explicit expressions for the matrix M and column vector V.

(c) For a generic data set, would you expect your system of equations to have a solution? *Briefly* explain your answer.

(d) Calculate $M^T M$ and $(M^T M)^{-1}$ (for the latter computation, state the condition required for the inverse to exist).

(e) Compute the least squares solution for m and b.

(f) The least squares method determines a vector X that minimizes the length of the vector $V - MX$. Draw a rough sketch of the three data points in the (x, y)-plane as well as their least squares fit. Indicate how the components of $V - MX$ could be obtained from your picture.

Solutions

1. You can find the definitions for all these terms by consulting the index of these notes.

2. Both junctions give the same equation for the currents

$$I + J + 13 = 0\,.$$

There are three voltage loops (one on the left, one on the right and one going around the outside of the circuit). Respectively, they give the equations

$$\begin{aligned} 60 - I - 80 - 3I &= 0 \\ 80 + 2J - V + 3J &= 0 \\ 60 - I + 2J - V + 3J - 3I &= 0 \quad . \end{aligned} \tag{2}$$

The above equations are easily solved (either using an augmented matrix and row reducing, or by substitution). The result is $I = -5$ Amps, $J = 8$ Amps, $V = 120$ Volts.

3. (a) m.
   (b) n.
   (c) Yes.
   (d) $n \times n$.
   (e) $m \times m$.
   (f) Yes. This relies on $\ker M = 0$ because if $M^T M$ had a non-trivial kernel, then there would be a non-zero solution X to $M^T M X = 0$. But then by multiplying on the left by X^T we see that $||MX|| = 0$. This in turn implies $MX = 0$ which contradicts the triviality of the kernel of M.
   (g) Yes because $\big(M^T M\big)^T = M^T (M^T)^T = M^T M$.
   (h) Yes, all symmetric matrices have a basis of eigenvectors.
   (i) No, because otherwise it would not be invertible.
   (j) Since the kernel of L is non-trivial, M must have 0 as an eigenvalue.

(k) Since M has a zero eigenvalue in this case, its determinant must vanish. I.e., $\det M = 0$.

4. To begin with the system becomes

$$\begin{pmatrix}1&1&1&1\\1&2&2&2\\1&2&3&3\end{pmatrix}\begin{pmatrix}x\\y\\z\\w\end{pmatrix}=\begin{pmatrix}1\\1\\1\end{pmatrix}$$

Then

$$M=\begin{pmatrix}1&1&1&1\\1&2&2&2\\1&2&3&3\end{pmatrix}=\begin{pmatrix}1&0&0\\1&1&0\\1&0&1\end{pmatrix}\begin{pmatrix}1&1&1&1\\0&1&1&1\\0&1&2&2\end{pmatrix}$$
$$=\begin{pmatrix}1&0&0\\1&1&0\\1&1&1\end{pmatrix}\begin{pmatrix}1&1&1&1\\0&1&1&1\\0&0&1&1\end{pmatrix}=LU$$

So now $MX=V$ becomes $LW=V$ where $W=UX=\begin{pmatrix}a\\b\\c\end{pmatrix}$ (say).

Thus we solve $LW=V$ by forward substitution

$$a=1,\ a+b=1,\ a+b+c=1\Rightarrow a=1,b=0,c=0\,.$$

Now solve $UX=W$ by back substitution

$$x+y+z+w=1,\ y+z+w=0,\ z+w=0$$

$$\Rightarrow w=\mu\ (\text{arbitrary}),z=-\mu,y=0,x=1\,.$$

The solution set is $\left\{\begin{pmatrix}x\\y\\z\\y\end{pmatrix}=\begin{pmatrix}1\\0\\-\mu\\\mu\end{pmatrix}:\mu\in\mathbb{R}\right\}$

5. ...

6. If U spans $\mathbb{R}^3$, then we must be able to express any vector $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ $\in \mathbb{R}^3$ as

$$X = c^1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + c^2 \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} + c^3 \begin{pmatrix} a \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & a \\ 0 & 2 & 1 \\ 1 & -3 & 0 \end{pmatrix} \begin{pmatrix} c^1 \\ c^2 \\ c^3 \end{pmatrix} ,$$

for some coefficients c^1, c^2 and c^3. This is a linear system. We could solve for c^1, c^2 and c^3 using an augmented matrix and row operations. However, since we know that $\dim \mathbb{R}^3 = 3$, if U spans $\mathbb{R}^3$, it will also be a basis. Then the solution for c^1, c^2 and c^3 would be unique. Hence, the 3×3 matrix above must be invertible, so we examine its determinant

$$\det \begin{pmatrix} 1 & 1 & a \\ 0 & 2 & 1 \\ 1 & -3 & 0 \end{pmatrix} = 1.(2.0 - 1.(-3)) + 1.(1.1 - a.2) = 4 - 2a .$$

Thus U spans $\mathbb{R}^3$ whenever $a \neq 2$. When $a = 2$ we can write the third vector in U in terms of the preceding ones as

$$\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \frac{3}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} .$$

(*You can obtain this result, or an equivalent one by studying the above linear system with $X = 0$, i.e., the associated homogeneous system.*)
The two vectors $\begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$ are clearly linearly independent, so this is the least number of vectors spanning U for this value of a. Also we see that $\dim U = 2$ in this case. Your picture should be a plane in $\mathbb{R}^3$ though the origin containing the vectors $\begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$.

7.
$$\det \begin{pmatrix} 1 & x \\ 1 & y \end{pmatrix} = y - x ,$$

$$\det\begin{pmatrix}1 & x & x^2\\ 1 & y & y^2\\ 1 & z & z^2\end{pmatrix} = \det\begin{pmatrix}1 & x & x^2\\ 0 & y-x & y^2-x^2\\ 0 & z-x & z^2-x^2\end{pmatrix}$$
$$= (y-x)(z^2-x^2)-(y^2-x^2)(z-x) = (y-x)(z-x)(z-y)\,.$$

$$\det\begin{pmatrix}1 & x & x^2 & x^3\\ 1 & y & y^2 & y^3\\ 1 & z & z^2 & z^3\\ 1 & w & w^2 & w^3\end{pmatrix} = \det\begin{pmatrix}1 & x & x^2 & x^3\\ 0 & y-x & y^2-x^2 & y^3-x^3\\ 0 & z-x & z^2-x^2 & z^3-x^3\\ 0 & w-x & w^2-x^2 & w^3-x^3\end{pmatrix}$$
$$= \det\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & y-x & y(y-x) & y^2(y-x)\\ 0 & z-x & z(z-x) & z^2(z-x)\\ 0 & w-x & w(w-x) & w^2(w-x)\end{pmatrix}$$
$$= (y-x)(z-x)(w-x)\det\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & y & y^2\\ 0 & 1 & z & z^2\\ 0 & 1 & w & w^2\end{pmatrix}$$
$$= (y-x)(z-x)(w-x)\det\begin{pmatrix}1 & x & x^2\\ 1 & y & y^2\\ 1 & z & z^2\end{pmatrix}$$
$$= (y-x)(z-x)(w-x)(y-x)(z-x)(z-y)\,.$$

From the 4×4 case above, you can see all the tricks required for a general Vandermonde matrix. First zero out the first column by subtracting the first row from all other rows (which leaves the determinant unchanged). Now zero out the top row by subtracting x_1 times the first column from the second column, x_1 times the second column from the third column *etc*. Again these column operations do not change the determinant. Now factor out x_2-x_1 from the second row, x_3-x_1 from the third row, *etc*. This does change the determinant so we write these factors outside the remaining determinant, which is just the same problem but for the $(n-1)\times(n-1)$ case. Iterating the same procedure gives the result

$$\det\begin{pmatrix}1 & x_1 & (x_1)^2 & \cdots & (x_1)^{n-1}\\ 1 & x_2 & (x_2)^2 & \cdots & (x_2)^{n-1}\\ 1 & x_3 & (x_3)^2 & \cdots & (x_3)^{n-1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & x_n & (x_n)^2 & \cdots & (x_n)^{n-1}\end{pmatrix} = \prod_{i>j}(x_i - x_j)\,.$$

(Here $\prod$ stands for a multiple product, just like Σ stands for a multiple sum.)

8. ...

9. (a)

$$\det\begin{pmatrix}\lambda & -\frac{1}{2} & -1\\ -\frac{1}{2} & \lambda-\frac{1}{2} & -\frac{1}{2}\\ -1 & -\frac{1}{2} & \lambda\end{pmatrix} = \lambda\Big((\lambda-\frac{1}{2}\Big)\lambda-\frac{1}{4})+\frac{1}{2}\Big(-\frac{\lambda}{2}-\frac{1}{2}\Big)-\Big(-\frac{1}{4}+\lambda\Big)$$

$$= \lambda^3 - \frac{1}{2}\lambda^2 - \frac{3}{2}\lambda = \lambda(\lambda+1)(\lambda-\frac{3}{2})\,.$$

Hence the eigenvalues are $0, -1, \frac{3}{2}$.

(b) When $\lambda = 0$ we must solve the homogenous system

$$\left(\begin{array}{ccc|c}0 & \frac{1}{2} & 1 & 0\\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0\\ 1 & \frac{1}{2} & 0 & 0\end{array}\right) \sim \left(\begin{array}{ccc|c}1 & \frac{1}{2} & 0 & 0\\ 0 & \frac{1}{4} & \frac{1}{2} & 0\\ 0 & \frac{1}{2} & 1 & 0\end{array}\right) \sim \left(\begin{array}{ccc|c}1 & 0 & -1 & 0\\ 0 & 1 & 2 & 0\\ 0 & 0 & 0 & 0\end{array}\right).$$

So we find the eigenvector $\begin{pmatrix}s\\ -2s\\ s\end{pmatrix}$ where $s \neq 0$ is arbitrary.

For $\lambda = -1$

$$\left(\begin{array}{ccc|c}1 & \frac{1}{2} & 1 & 0\\ \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & 0\\ 1 & \frac{1}{2} & 1 & 0\end{array}\right) \sim \left(\begin{array}{ccc|c}1 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\end{array}\right).$$

So we find the eigenvector $\begin{pmatrix} -s \\ 0 \\ s \end{pmatrix}$ where $s \neq 0$ is arbitrary.

Finally, for $\lambda = \frac{3}{2}$

$$\left(\begin{array}{ccc|c} -\frac{3}{2} & \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & -\frac{3}{2} & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & \frac{1}{2} & -\frac{3}{2} & 0 \\ 0 & -\frac{5}{4} & \frac{5}{4} & 0 \\ 0 & \frac{5}{4} & -\frac{5}{4} & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

So we find the eigenvector $\begin{pmatrix} s \\ s \\ s \end{pmatrix}$ where $s \neq 0$ is arbitrary.

If the mistake X is in the direction of the eigenvector $\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$, then $Y = 0$. *I.e.*, the satellite returns to the origin $\mathcal{O}$. For all subsequent orbits it will again return to the origin. NASA would be very pleased in this case.

If the mistake X is in the direction $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$, then $Y = -X$. Hence the satellite will move to the point opposite to X. After next orbit will move back to X. It will continue this wobbling motion indefinitely. Since this is a stable situation, again, the elite engineers will pat themselves on the back.

Finally, if the mistake X is in the direction $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, the satellite will move to a point $Y = \frac{3}{2}X$ which is further away from the origin. The same will happen for all subsequent orbits, with the satellite moving a factor 3/2 further away from $\mathcal{O}$ each orbit (in reality, after several orbits, the approximations used by the engineers in their calculations probably fail and a new computation will be needed). In this case, the satellite will be lost in outer space and the engineers will likely lose their jobs!

10. (a) A basis for B^3 is $\left\{ \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix} \right\}$

(b) 3.

(c) $2^3 = 8$.

(d) $\dim B^3 = 3$.

(e) Because the vectors $\{v_1, v_2, v_3\}$ are a basis any element $v \in B^3$ can be written uniquely as $v = b^1 v_1 + b^2 v_2 + b^3 v_3$ for some triplet of bits $\begin{pmatrix}b^1\\b^2\\b^3\end{pmatrix}$. Hence, to compute $L(v)$ we use linearity of L

$$L(v) = L(b^1 v_1 + b^2 v_2 + b^3 v_3) = b^1 L(v_1) + b^2 L(v_2) + b^3 L(v_3)$$
$$= \begin{pmatrix} L(v_1) & L(v_2) & L(v_3) \end{pmatrix} \begin{pmatrix}b^1\\b^2\\b^3\end{pmatrix}.$$

(f) From the notation of the previous part, we see that we can list linear transformations $L : B^3 \to B$ by writing out all possible bit-valued row vectors

$$\begin{pmatrix}0 & 0 & 0\end{pmatrix},$$
$$\begin{pmatrix}1 & 0 & 0\end{pmatrix},$$
$$\begin{pmatrix}0 & 1 & 0\end{pmatrix},$$
$$\begin{pmatrix}0 & 0 & 1\end{pmatrix},$$
$$\begin{pmatrix}1 & 1 & 0\end{pmatrix},$$
$$\begin{pmatrix}1 & 0 & 1\end{pmatrix},$$
$$\begin{pmatrix}0 & 1 & 1\end{pmatrix},$$
$$\begin{pmatrix}1 & 1 & 1\end{pmatrix}.$$

There are $2^3 = 8$ different linear transformations $L : B^3 \to B$, exactly the same as the number of elements in B^3.

(g) Yes, essentially just because L_1 and L_2 are linear transformations. In detail for any bits (a, b) and vectors (u, v) in B^3 it is easy to check the linearity property for $(\alpha L_1 + \beta L_2)$

$$(\alpha L_1 + \beta L_2)(au + bv) = \alpha L_1(au + bv) + \beta L_2(au + bv)$$

$$
\begin{aligned}
&= \alpha a L_1(u) + \alpha b L_1(v) + \beta a L_1(u) + \beta b L_1(v) \\
&= a(\alpha L_1(u) + \beta L_2(v)) + b(\alpha L_1(u) + \beta L_2(v)) \\
&= a(\alpha L_1 + \beta L_2)(u) + b(\alpha L_1 + \beta L_2)(v)\,.
\end{aligned}
$$

Here the first line used the definition of $(\alpha L_1 + \beta L_2)$, the second line depended on the linearity of L_1 and L_2, the third line was just algebra and the fourth used the definition of $(\alpha L_1 + \beta L_2)$ again.

(h) Yes. The easiest way to see this is the identification above of these maps with bit-valued column vectors. In that notation, a basis is

$$
\left\{ \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \right\}.
$$

Since this (spanning) set has three (linearly independent) elements, the vector space of linear maps $B^3 \to B$ has dimension 3. This is an example of a general notion called the *dual vector space*.

11. ...

12. (a) If we call $M = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$, then $X^T M X = ax^2 + 2bxy + dy^2$. Similarly putting $C = \begin{pmatrix} c \\ e \end{pmatrix}$ yields $X^T C + C^T X = 2X \cdot C = 2cx + 2ey$. Thus

$$
\begin{aligned}
0 &= ax^2 + 2bxy + dy^2 + 2cx + 2ey + f \\
&= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ b & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} c \\ e \end{pmatrix} + \begin{pmatrix} c & e \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + f\,.
\end{aligned}
$$

(b) Yes, the matrix M is symmetric, so it will have a basis of eigenvectors and is similar to a diagonal matrix of real eigenvalues.

To find the eigenvalues notice that $\det \begin{pmatrix} a-\lambda & b \\ b & d-\lambda \end{pmatrix} = (a - \lambda)(d-\lambda) - b^2 = \left(\lambda - \frac{a+d}{2}\right)^2 - b^2 - \left(\frac{a-d}{2}\right)^2$. So the eigenvalues are

$$
\lambda = \frac{a+d}{2} + \sqrt{b^2 + \left(\frac{a-d}{2}\right)^2} \text{ and } \mu = \frac{a+d}{2} - \sqrt{b^2 + \left(\frac{a-d}{2}\right)^2}\,.
$$

(c) The trick is to write

$$X^TMX+C^TX+X^TC = (X^T+C^TM^{-1})M(X+M^{-1}C)-C^TM^{-1}C\,,$$

so that

$$(X^T + C^TM^{-1})M(X + M^{-1}C) = C^TMC - f\,.$$

Hence $Y = X + M^{-1}C$ and $g = C^TMC - f$.

(d) The cosine of the angle between vectors V and W is given by

$$\frac{V \cdot W}{\sqrt{V \cdot V\, W \cdot W}} = \frac{V^TW}{\sqrt{V^TV\, W^TW}}\,.$$

So replacing $V \to PV$ and $W \to PW$ will always give a factor P^TP inside all the products, but $P^TP = I$ for orthogonal matrices. Hence none of the dot products in the above formula changes, so neither does the angle between V and W.

(e) If we take the eigenvectors of M, normalize them (*i.e.* divide them by their lengths), and put them in a matrix P (as columns) then P will be an orthogonal matrix. (If it happens that $\lambda = \mu$, then we also need to make sure the eigenvectors spanning the two dimensional eigenspace corresponding to λ are orthogonal.) Then, since M times the eigenvectors yields just the eigenvectors back again multiplied by their eigenvalues, it follows that $MP = PD$ where D is the diagonal matrix made from eigenvalues.

(f) If $Y = PZ$, then $Y^TMY = Z^TP^TMPZ = Z^TP^TPDZ = Z^TDZ$ where $D = \begin{pmatrix}\lambda & 0\\ 0 & \mu\end{pmatrix}$.

(g) Using part (f) and (c) we have

$$\lambda z^2 + \mu w^2 = g\,.$$

(h) When $\lambda = \mu$ and $g/\lambda = R^2$, we get the equation for a circle radius R in the (z, w)-plane. When λ, μ and g are postive, we have the equation for an ellipse. Vanishing g along with λ and μ of opposite signs gives a pair of straight lines. When g is non-vanishing, but λ and μ have opposite signs, the result is a pair of hyperbolæ. These shapes all come from cutting a cone with a plane, and are therefore called conic sections.

13. We show that L is bijective if and only if M is invertible.

(a) We suppose that L is bijective.

i. Since L is injective, its kernel consists of the zero vector alone. Hence

$$L = \dim \ker L = 0.$$

So by the Dimension Formula,

$$\dim V = L + \operatorname{rank} L = \operatorname{rank} L.$$

Since L is surjective, $L(V) = W$. Thus

$$\operatorname{rank} L = \dim L(V) = \dim W.$$

Thereby

$$\dim V = \operatorname{rank} L = \dim W.$$

ii. Since $\dim V = \dim W$, the matrix M is square so we can talk about its eigenvalues. Since L is injective, its kernel is the zero vector alone. That is, the only solution to $LX = 0$ is $X = 0_V$. But LX is the same as MX, so the only solution to $MX = 0$ is $X = 0_V$. So M does not have zero as an eigenvalue.

iii. Since $MX = 0$ has no non-zero solutions, the matrix M is invertible.

(b) Now we suppose that M is an invertible matrix.

i. Since M is invertible, the system $MX = 0$ has no non-zero solutions. But LX is the same as MX, so the only solution to $LX = 0$ is $X = 0_V$. So L does not have zero as an eigenvalue.

ii. Since $LX = 0$ has no non-zero solutions, the kernel of L is the zero vector alone. So L is injective.

iii. Since M is invertible, we must have that $\dim V = \dim W$. By the Dimension Formula, we have

$$\dim V = L + \operatorname{rank} L$$

and since $\ker L = \{0_V\}$ we have $L = \dim \ker L = 0$, so

$$\dim W = \dim V = \operatorname{rank} L = \dim L(V).$$

Since $L(V)$ is a subspace of W with the same dimension as W, it must be equal to W. To see why, pick a basis B of $L(V)$. Each element of B is a vector in W, so the elements of B form a linearly independent set in W. Therefore B is a basis of W, since the size of B is equal to $\dim W$. So $L(V) = \operatorname{span} B = W$. So L is surjective.

14. (a) $F_4 = F_2 + F_3 = 2 + 3 = 5$.

(b) The number of pairs of doves in any given year equals the number of the previous years plus those that hatch and there are as many of them as pairs of doves in the year before the previous year.

(c) $X_1 = \begin{pmatrix} F_1 \\ F_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $X_2 = \begin{pmatrix} F_2 \\ F_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

$$MX_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = X_2\,.$$

(d) We just need to use the recursion relationship of part (b) in the top slot of X_{n+1}:

$$X_{n+1} = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = \begin{pmatrix} F_n + F_{n-1} \\ F_n \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix} = MX_n\,.$$

(e) Notice M is symmetric so this is guaranteed to work.

$$\det \begin{pmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{pmatrix} = \lambda(\lambda - 1) - 1 = \big(\lambda - \frac{1}{2}\big)^2 - \frac{5}{4}\,,$$

so the eigenvalues are $\frac{1\pm\sqrt{5}}{2}$. Hence the eigenvectors are $\begin{pmatrix} \frac{1\pm\sqrt{5}}{2} \\ 1 \end{pmatrix}$, respectively (notice that $\frac{1+\sqrt{5}}{2} + 1 = \frac{1+\sqrt{5}}{2} . \frac{1+\sqrt{5}}{2}$ and $\frac{1-\sqrt{5}}{2} + 1 = \frac{1-\sqrt{5}}{2} . \frac{1-\sqrt{5}}{2}$). Thus $M = PDP^{-1}$ with

$$D = \begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix} \text{ and } P = \begin{pmatrix} \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix}.$$

(f) $M^n = (PDP^{-1})^n = PDP^{-1}PDP^{-1} \ldots PDP^{-1} = PD^nP^{-1}$.

(g) Just use the matrix recursion relation of part (d) repeatedly:

$$X_{n+1} = MX_n = M^2 X_{n-1} = \cdots = M^n X_1 \, .$$

(h) The eigenvalues are $\varphi = \frac{1+\sqrt{5}}{2}$ and $1 - \varphi = \frac{1-\sqrt{5}}{2}$.

(i)

$$X_{n+1} = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = M^n X_n = PD^n P^{-1} X_1$$

$$= P \begin{pmatrix} \varphi & 0 \\ 0 & 1-\varphi \end{pmatrix}^n \begin{pmatrix} \frac{1}{\sqrt{5}} & \star \\ -\frac{1}{\sqrt{5}} & \star \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = P \begin{pmatrix} \varphi^n & 0 \\ 0 & (1-\varphi)^n \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{\varphi^n}{\sqrt{5}} \\ -\frac{(1-\varphi)^n}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \star \\ \frac{\varphi^n - (1-\varphi)^n}{\sqrt{5}} \end{pmatrix} .$$

Hence

$$F_n = \frac{\varphi^n - (1-\varphi)^n}{\sqrt{5}} \, .$$

These are the famous Fibonacci numbers.

15. Call the three vectors u, v and w, respectively. Then

$$v^{\perp} = v - \frac{u \bullet v}{u \bullet u} u = v - \frac{3}{4} u = \begin{pmatrix} \frac{1}{4} \\ -\frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix} ,$$

and

$$w^{\perp} = w - \frac{u \bullet w}{u \bullet u} u - \frac{v^{\perp} \bullet w}{v^{\perp} \bullet v^{\perp}} v^{\perp} = w - \frac{3}{4} u - \frac{\frac{3}{4}}{\frac{3}{4}} v^{\perp} = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Dividing by lengths, an orthonormal basis for $\text{span}\{u, v, w\}$ is

$$\left\{ \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{\sqrt{3}}{6} \\ -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{6} \\ \frac{\sqrt{3}}{6} \end{pmatrix}, \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ 0 \\ 0 \\ \frac{\sqrt{2}}{2} \end{pmatrix} \right\} .$$

16. ...

17. ...

18. We show that L is bijective if and only if M is invertible.

   (a) We suppose that L is bijective.

      i. Since L is injective, its kernel consists of the zero vector alone. So
      $$L = \dim \ker L = 0.$$
      So by the Dimension Formula,
      $$\dim V = L + \operatorname{rank} L = \operatorname{rank} L.$$
      Since L is surjective, $L(V) = W$. So
      $$\operatorname{rank} L = \dim L(V) = \dim W.$$
      So
      $$\dim V = \operatorname{rank} L = \dim W.$$

      ii. Since $\dim V = \dim W$, the matrix M is square so we can talk about its eigenvalues. Since L is injective, its kernel is the zero vector alone. That is, the only solution to $LX = 0$ is $X = 0_V$. But LX is the same as MX, so the only solution to $MX = 0$ is $X = 0_V$. So M does not have zero as an eigenvalue.

      iii. Since $MX = 0$ has no non-zero solutions, the matrix M is invertible.

   (b) Now we suppose that M is an invertible matrix.

      i. Since M is invertible, the system $MX = 0$ has no non-zero solutions. But LX is the same as MX, so the only solution to $LX = 0$ is $X = 0_V$. So L does not have zero as an eigenvalue.

      ii. Since $LX = 0$ has no non-zero solutions, the kernel of L is the zero vector alone. So L is injective.

      iii. Since M is invertible, we must have that $\dim V = \dim W$. By the Dimension Formula, we have
      $$\dim V = L + \operatorname{rank} L$$

and since $\ker L = \{0_V\}$ we have $L = \dim \ker L = 0$, so

$$\dim W = \dim V = \operatorname{rank} L = \dim L(V).$$

Since $L(V)$ is a subspace of W with the same dimension as W, it must be equal to W. To see why, pick a basis B of $L(V)$. Each element of B is a vector in W, so the elements of B form a linearly independent set in W. Therefore B is a basis of W, since the size of B is equal to $\dim W$. So $L(V) = \operatorname{span} B = W$. So L is surjective.

19. ...

Index

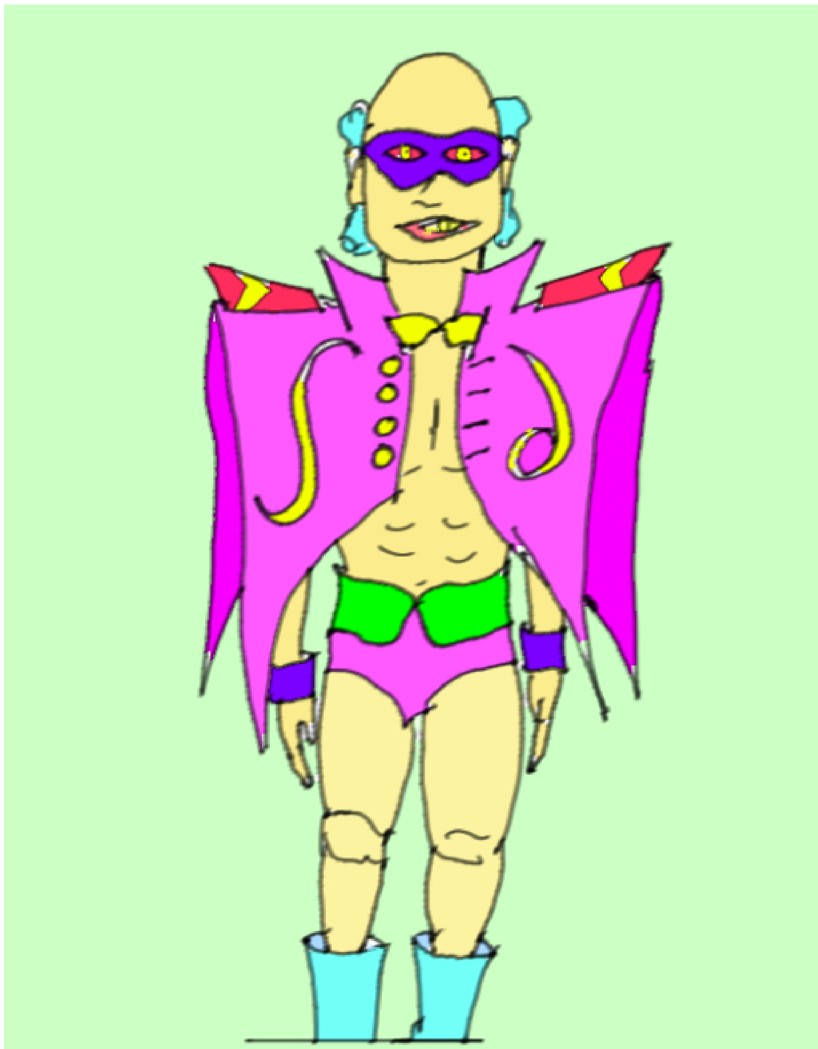


Figure 2: Captain Conundrum