Linear Algebra in Twenty Five Lectures

Tom Denton and Andrew Waldron

December 12, 2009

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 teach basic linear algebra skills. These are a set of nine Webwork assignments which are collected weekly and available at

http://webwork.math.ucdavis.edu/

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 a 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 textbook of Jim Hefferon:

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

and the notes are also hyperlinked to Wikipedia where students can rapidly access further details and background material for many of the concepts. There are also an array of useful commercially available texts such as

- "Introductory Linear Algebra, An Applied First Course", B. Kolman and D. Hill, Pearson 2001.
- "Algebra and Geometry", D. Holten and J. Lloyd, CBRC, 1978.
- "Theory and Problems of Linear Algebra", S. Lipschutz, McGraw-Hill 1987.

There are still many errors in the notes, as well as awkwardly explained concepts. An army of 200 students have already found many of them. 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.

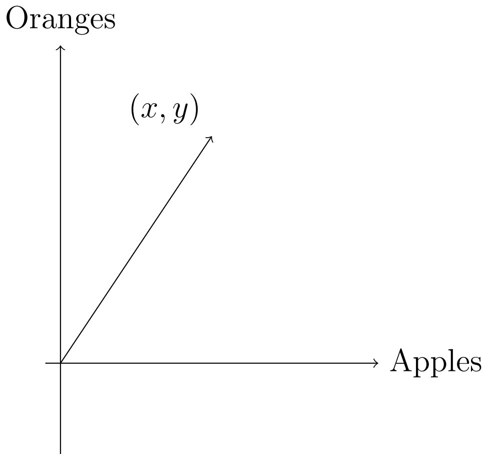
1 What is Linear Algebra?

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, or any places where x and y are multiplied together.

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:

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

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:

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

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 linear systems, we'll learn a general method called Gaussian Elimination.

References

Hefferon, Chapter One, Section 1
  Wikipedia, Systems of Linear Equations

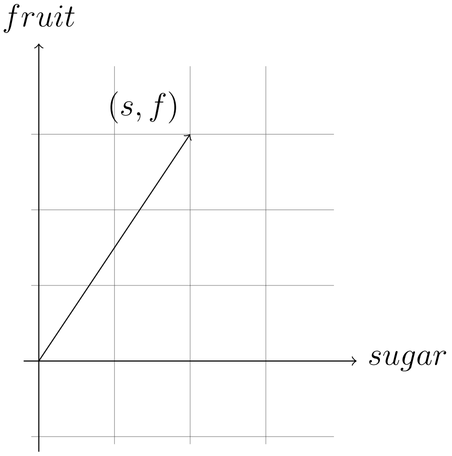
Review Problems

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. 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, the 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.

3. 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

Last time 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:

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

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 for any vector v, 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.

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

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: $x = 9, y = 18$.

Generally, we say that two augmented matrices are (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:

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

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 $z = 4 - 7\lambda$. 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.

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

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

Next time, we'll prove it.

References

Hefferon, Chapter One, Section 1
Wikipedia, Row Echelon Form

Review Problems

1. Show that this pair of augmented matrices are row equivalent, assuming $ad - bc \neq 0$.
   $\left(\begin{array}{cc|c} a & b & x \\ c & d & y \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & \frac{dx-by}{ad-bc} \\ 0 & 1 & \frac{ay-cx}{ad-bc} \end{array}\right)$

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

   Give a *geometric* reason that the associated system of equations has no solution. Given a general augmented matrix $\left(\begin{array}{cc|c} a & b & x \\ c & d & y \end{array}\right)$, can you find a condition on the numbers a, b, c and d that create the geometric condition you found?

3. 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.

4. Row equivalence of matrices is an example of an *equivalence relation.* 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 $x \sim z$.

   (a) Consider the real numbers with the relation $\geq$. Is this an equivalence relation? Why or why not?

   (b) Consider the set of Euclidean triangles with the relation of similarity. (Recall that two triangles are similar if all of their angles are equal.) Is this an equivalence relation? Why or why not?

(c) 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 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.

Example

$$\begin{array}{rrrl} & & 3x_3 & = 9 \\ x_1 & +5x_2 & -2x_3 & = 2 \\ \frac{1}{3}x_1 & +2x_2 & & = 3 \end{array}$$

First we write the system as an augmented matrix:

$$
\begin{aligned}
\left(\begin{array}{ccc|c} 0 & 0 & 3 & 9 \\ 1 & 5 & -2 & 2 \\ \frac{1}{3} & 2 & 0 & 3 \end{array}\right)
\quad & \overset{R_1 \leftrightarrow R_2}{\sim} \quad
\left(\begin{array}{ccc|c} \frac{1}{3} & 2 & 0 & 3 \\ 1 & 5 & -2 & 2 \\ 0 & 0 & 3 & 9 \end{array}\right) \\
& \overset{3R_1}{\sim} \quad
\left(\begin{array}{ccc|c} 1 & 6 & 0 & 9 \\ 1 & 5 & -2 & 2 \\ 0 & 0 & 3 & 9 \end{array}\right) \\
& \overset{R_2 = R_2 - R_1}{\sim} \quad
\left(\begin{array}{ccc|c} 1 & 6 & 0 & 9 \\ 0 & -1 & -2 & -7 \\ 0 & 0 & 3 & 9 \end{array}\right) \\
& \overset{-R_2}{\sim} \quad
\left(\begin{array}{ccc|c} 1 & 6 & 0 & 9 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 3 & 9 \end{array}\right) \\
& \overset{R_1 = R_1 - 6R_2}{\sim} \quad
\left(\begin{array}{ccc|c} 1 & 0 & -12 & -33 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 3 & 9 \end{array}\right) \\
& \overset{\frac{1}{3}R_3}{\sim} \quad
\left(\begin{array}{ccc|c} 1 & 0 & -12 & -33 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 1 & 3 \end{array}\right) \\
& \overset{R_1 = R_1 + 12R_3}{\sim} \quad
\left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 1 & 3 \end{array}\right) \\
& \overset{R_2 = R_2 - 2R_3}{\sim} \quad
\left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{array}\right)
\end{aligned}
$$

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

$$
\left(\begin{array}{cccc|c}
1 & 0 & -1 & 2 & -1 \\
1 & 1 & 1 & -1 & 2 \\
0 & -1 & -2 & 3 & -3 \\
5 & 2 & -1 & 4 & 1
\end{array}\right)
$$

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

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

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:

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

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 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 the exercise, removing columns does not affect row equivalence. Call the new, smaller, matrices $\hat{A}$ and $\hat{B}$. The new matrices should look this: $\hat{A} = \left(\begin{array}{c|c} I_N & a \\ 0 & 0 \end{array}\right)$ 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 $A = B$. □

References

Hefferon, Chapter One, Section 1.1 and 1.2

Wikipedia, Row Echelon Form

Wikipedia, Elementary Matrix Operations

Review Problems

1. Explain why row equivalence is not affected by removing columns. Is row equivalence affected by removing rows? Prove or give a counterexample.

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?

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

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 the sake of visualization, consider the case of r equations in three variables. Geometrically, then, each of our 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.

4.1 Non-Leading Variables

Variables that are not a pivot in the reduced row echelon form of a linear system are *free*. Se 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:

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

It's worth noting that if we knew how to multiply matrices of any size, we could write the previous system as $MX = V$, where

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

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

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

Definition Let X and Y by 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.

Soon we'll show that matrix multiplication is 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:

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

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

$MY_2 = 0$ says that $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$ 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:

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

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

References

Hefferon, Chapter One, Section I.2

Wikipedia, Systems of Linear Equations

Review Questions

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

2. Let

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

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

$$\begin{array}{cccc} a_1^1 x^1 & +a_2^1 x^2 & \dots +a_k^1 x^k & = 0 \\ a_1^2 x^1 & +a_2^2 x^2 & \dots +a_k^2 x^k & = 0 \\ \vdots & \vdots & \vdots & \vdots \\ a_1^r x^1 & +a_2^r x^2 & \dots +a_k^r x^k & = 0 \end{array}$$

   Does your rule for multiplying a matrix times a vector obey the linearity property? Prove it!

3. 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 the last problem, find a simple rule for multiplying Me_i, where M is the general matrix defined in the last problem.

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[1].

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

Definition Given two vectors, $u = \begin{pmatrix} u^1 \\ \vdots \\ u^n \end{pmatrix}$ and $v = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}$ their *sum*

$$u + v = \begin{pmatrix} u^1 + v^1 \\ \vdots \\ u^n + v^n \end{pmatrix} .$$

Given a scalar c, the *scalar multiple*

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

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

[1] 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!

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, $\{\begin{pmatrix}1\\2\\3\\4\end{pmatrix} + t\begin{pmatrix}1\\0\\0\\0\end{pmatrix} | t \in \mathbb{R}\}$ 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 $\cos\theta$:

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

Thus,

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

This 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

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

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

There are, in fact, many different useful ways to define lengths of vectors. Notice in the definition above how we defined the dot product, and then all the other definitions are dependent on the definition 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 provide the *Euclidean length and angle* between two vectors.

Other definitions of length and angle arise from *bilinear forms*, which have all of the properties listed above. Instead of writing $\cdot$ for other bilinear forms, we usually write $\langle u, v\rangle$ to avoid confusion.

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 "length" of a vector with coordinates x, y, z and t is $||v|| = x^2 + y^2 + z^2 - t^2$.

Theorem 5.1 (Cauchy-Schwartz Inequality)**.** *For vectors u and v with an inner-product $\langle\,,\rangle$,*
$$\frac{|\langle u, v\rangle|}{||u||\,||v||} \leq 1$$

Proof. This follows from the definition of the angle between two vectors and the fact that $\cos\theta \leq 1$. □

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. □

References

Hefferon: Chapter One.II

Relevant Wikipedia Articles:

- Dot Product
- Inner Product Space
- Minkoski Metric

Review Questions

1. (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?

2. 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 $\mathbb{R}^2$.

   (b) Compute $\frac{||MX||}{||X||}$.

3. 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.

4. (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 idea of a vector can be much more general. In the spirit of generalization, then, we will 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 worl 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) $u + v \in \mathbb{R}$. (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 $\cdot : \mathbb{R} \times V \to V$.) On the other hand, the dot product takes two vectors and returns a number. (In notation: $\bullet : V \times V \to \mathbb{R}$.)

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

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

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 column vectors: $f(0)$ is the first entry, $f(1)$ is the second entry, and so on. Then for example the function $f(n) = n^3$ would look like this:

$$f(n) = \begin{pmatrix} 0 \\ 1 \\ 8 \\ 27 \\ \vdots \\ n^3 \\ \vdots \end{pmatrix}$$

Alternatively, V is the space of sequences: $f = \{f_1, f_2, \ldots, f_n, \ldots\}$.

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.

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

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

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, $\mathbb{C} = \{x + iy | i^2 = -1, x, y \in \mathbb{R}\}$. In quantum physics, vector spaces over $\mathbb{C}$ describe all possible states a system of particles can have.

For example,

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

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 $4\cdot 4=1$. 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

Wikipedia:

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

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

2. Consider the set of convergent sequences, with the same addition and scalar multiplication that we defined for the space of sequences:

   $$V = \{f | f : \mathbb{N} \to \mathbb{R}, \lim_{n\to\infty} f \in \mathbb{R}\}$$

   Is this still a vector space? Explain why or why not.

3. Let $V = \{\begin{pmatrix} x \\ y \end{pmatrix} : x, y \in \mathbb{R}\} = \mathbb{R}^2$.

   Propose as many rules for addition and scalar multiplication as you can that satisfy some of the vector space conditions while breaking some others.

4. Consider the set of 2×4 matrices:

   $$V = \{\begin{pmatrix} a & b & c & d \\ e & f & g & h \end{pmatrix} | a, b, c, d, e, f, g, h \in \mathbb{C}\}$$

   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:

$$L : V \to W$$

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 $L : V \to W$ 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.

Example Take $L : \mathbb{R}^3 \to \mathbb{R}^3$ 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.

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

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 L using a matrix like so:

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

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.

$$V = \{a_0 + a_1 x + \ldots + a_n x^n | n \in \mathbb{N}, a_i \in \mathbb{R}\}$$

Let $L : V \to V$ 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:

$$a_0 + a_1 x + \ldots + a_n x^n \longleftrightarrow \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

Then we have:

$$\frac{d}{dx}(a_0 + a_1 x + \ldots + a_n x^n) = a_1 + 2a_2 x + \ldots + na_n x^{n-1} \longleftrightarrow \begin{pmatrix} a_1 \\ 2a_2 \\ \vdots \\ na_n \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

One could then write the derivative as an infinite matrix:

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

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.)

Wikipedia:

- Linear Transformation
- Dimension

Review Questions

1. Show that the pair of conditions:

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

is equivalent to the single condition:

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

Your answer should have two parts. Show that $(1, 2) \Rightarrow (3)$, and then show that $(3) \Rightarrow (1, 2)$.

2. Let P_n be the space of degree n polynomials 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?

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.

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 is (perhaps unsurprisingly) called a *row vector*.

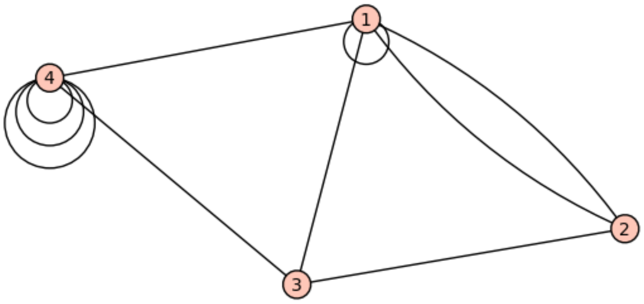
Matrices are a very useful and efficient way to store information:

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.

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

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

Here we multiply an $r \times k$ matrix by a $k \times 1$ vector to produce a $r \times 1$ vector.

Likewise, we can use matrices $N = (n^i_j)$ to represent linear transformations

$$M^s_k \xrightarrow{N} M^r_k$$

via $(L(M)^i_l) = (\sum_{j=1}^k n^i_j m^j_l)$. 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:

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

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

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

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 $\mathbb{1}$. The identity matrix is spacial 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 $\begin{pmatrix}2 & 5 & 6\\1 & 3 & 4\end{pmatrix}^T = \begin{pmatrix}2 & 1\\5 & 3\\6 & 4\end{pmatrix}$

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 $(MN)^T = N^T M^T$.*

References

Hefferon, Chapter Three, Section IV, parts 1-3.

Wikipedia:

- Matrix Multiplication

Review Questions

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

2. Prove the theorem $(MN)^T = N^T M^T$.

3. Let M be any $m \times n$ matrix. Show that $M^T M$ is a symmetric $n \times n$ matrix.

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. Prove that the dot product $x \cdot y = x^T \; \mathbb{1} \; y$.

5. Explain what happens to a matrix when:

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

   (ii) 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.

- 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:

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

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

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

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 $\exp(M) = 1+M+\frac{1}{2}M^2+\frac{1}{3!}M^3+\dots$ 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, $\mathrm{tr}(MN) = 2 + 1 = 3 = 1 + 2 = \mathrm{tr}(NM)$.

Another useful property of the trace is that:

$$\mathrm{tr}\, M = \mathrm{tr}\, M^T$$

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

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

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 $(AB)^T = B^TA^T$. Since $I^T = I$, then $(A^{-1}A)^T = A^T(A^{-1})^T = I$. Similarly, $(AA^{-1})^T = (A^{-1})^TA^T = I$. 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} 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.

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

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 $M^{-1}M = I$):

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

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

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:

$$\begin{aligned} -x \, +2y \, -3z &= 1 \\ 2x \, +y \qquad &= 2 \\ 4x \, -2y \, +5z &= 0 \end{aligned}$$

The associated matrix equation is $MX = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$. Then:

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

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

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

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. □

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:

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

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 very difficult to break cryptographic systems. 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 substutions 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 $\mathbb{Z}_2^8$. 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.

References

Hefferon: Chapter Three, Section IV.2

Wikipedia: Invertible Matrix

Review Questions

1. 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.

   (Show that $(i) \Rightarrow (ii)$ and $(ii) \Rightarrow (i)$.)

2. 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?

3. 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.

11 LU Decomposition

Certain matrices are easier to work with than others. In this section, we will see how to write any square[2] matrix M as the product of two simpler matrices. We'll 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 triangular matrices are very easy to solve by back substitution.

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

[2] 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.

Example Consider the linear system:

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

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 $LW = V$:

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

  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 $x = -6$.

  Then $X = \begin{pmatrix} -6 \\ 3 \\ -11 \end{pmatrix}$, 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.

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:

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

   i. Find x^1.

ii. Find x^2.

iii. Find x^3.

k. Try to find a formula for x^k.

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 is invertible, 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 $M = \begin{pmatrix} m_1^1 & m_2^1 \\ m_1^2 & m_2^2 \end{pmatrix}$, then $M^{-1} = \frac{1}{m_1^1 m_2^2 - m_2^1 m_1^2} \begin{pmatrix} m_2^2 & -m_2^1 \\ -m_1^2 & m_1^1 \end{pmatrix}$. Thus M is invertible if and only if

$$m_1^1 m_2^2 - m_2^1 m_1^2 \neq 0\,.$$

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

$$\det M = \det \begin{pmatrix} m_1^1 & m_2^1 \\ m_1^2 & m_2^2 \end{pmatrix} = m_1^1 m_2^2 - m_2^1 m_1^2$$

Example For a 3×3 matrix, $M = \begin{pmatrix} m_1^1 & m_2^1 & m_3^1 \\ m_1^2 & m_2^2 & m_3^2 \\ m_1^3 & m_2^3 & m_3^3 \end{pmatrix}$, then (by the first review question) M is non-singular if and only if:

$$\det M = m_1^1 m_2^2 m_3^3 - m_1^1 m_3^2 m_2^3 + m_2^1 m_3^2 m_1^3 - m_2^1 m_1^2 m_3^3 + m_3^1 m_1^2 m_2^3 - m_3^1 m_2^2 m_1^3 \neq 0.$$

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:

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

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:

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

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:

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

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:

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

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 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:

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

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

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

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
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. *i.* 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?

   ii. 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.

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

3. Let M a matrix and E^i_j the elementary matrix swapping two rows. Explain every line of the series of equations proving that $\det M = -\det(E^i_j M)$.

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.

   i. 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? What is $\tau^2_{i,j}$?

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

iii. 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.)*

13 Elementary Matrices and Determinants II

In the last section, 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?

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

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.

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

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 an 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* $\det(EM) = \det E \det M$.

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 $\operatorname{RREF}(M)$ is the reduced row echelon form of M. Then $\operatorname{RREF}(M) = E_1 E_2 \dots E_k M$ 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 $\operatorname{RREF}(M)$ 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, $\det \operatorname{RREF}(M) = \lambda \det \operatorname{RREF}(M)$, and so $\det \operatorname{RREF}(M) = 0$.

- 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 $\det \mathrm{RREF}(M) = \det(E_1E_2 \ldots E_kM)$. Then by the theorem above, $\det \mathrm{RREF}(M) = \det(E_1) \ldots \det(E_k) \det M$. 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,

$$M = E_1E_2 \ldots E_k \ \mathrm{RREF}(M)\,,$$

and

$$N = F_1F_2 \ldots F_l \ \mathrm{RREF}(N) = N\,.$$

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

$$\begin{aligned}
\det(MN) &= \det(E_1E_2 \ldots E_k \ \mathrm{RREF}(M) F_1F_2 \ldots F_l \ \mathrm{RREF}(N)) \\
&= \det(E_1E_2 \ldots E_k I F_1F_2 \ldots F_l \ \mathrm{RREF}(N)) \\
&= \det(E_1) \ldots \det(E_k) \det(I) \det(F_1) \ldots \det(F_l) \det(\mathrm{RREF}(N) \\
&= \det(M) \det(N)
\end{aligned}$$

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

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

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!

References

Hefferon, Chapter Four, Section I.1 and I.3

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:

   - $\det M$.
   - $\det N$.
   - $\det(MN)$.
   - $\det M \det N$.
   - $\det(M^{-1})$ assuming $ab - cd \neq 0$.

- $\det(M^T)$
- $\det(M+N) - \det M - \det N$

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

Last time 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:

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

A *minor* of an $n \times n$ matrix M is 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 the determinants of its minors as follows:

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

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 $\{2, \dots, n\} \to \{1, \dots, j-1, j+1, \dots, n\}$. We call this partial permutation $\hat\sigma = \begin{bmatrix}\sigma(2) & \dots & \sigma(n)\end{bmatrix}$.

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 $\begin{bmatrix}j & \sigma(2) & \dots & \sigma(n)\end{bmatrix}$, 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[3] $\det M = 0$

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:

$$\begin{aligned}\det\begin{pmatrix}1 & 2 & 3\\ 4 & 0 & 0\\ 7 & 8 & 9\end{pmatrix} &= -\det\begin{pmatrix}4 & 0 & 0\\ 1 & 2 & 3\\ 7 & 8 & 9\end{pmatrix}\\ &= 4\det\begin{pmatrix}2 & 3\\ 8 & 9\end{pmatrix}\\ &= -16\end{aligned}$$

Theorem 14.1. *For any square matrix M, we have:*

$$\det M^T = \det M$$

[3] A fun exercise is to compute the determinant of a 4 × 4 matrix filled in order, from left to right, with the numbers $1, 2, 3, \ldots 16$. What do you observe? Try the same for a 5 × 5 matrix with $1, 2, 3 \ldots 25$. Is there a pattern? Can you explain it?

Proof. By definition,

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

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, $\operatorname{sgn} \sigma = \operatorname{sgn} \sigma^{-1}$.

Then we can write out the above in formulas as follows:

$$\begin{aligned}
\det M &= \sum_{\sigma} \operatorname{sgn}(\sigma) m^1_{\sigma(1)} m^2_{\sigma(2)} \dots m^n_{\sigma(n)} \\
&= \sum_{\sigma} \operatorname{sgn}(\sigma) m_1^{\sigma^{-1}(1)} m_2^{\sigma^{-1}(2)} \dots m_n^{\sigma^{-1}(n)} \\
&= \sum_{\sigma} \operatorname{sgn}(\sigma^{-1}) m_1^{\sigma^{-1}(1)} m_2^{\sigma^{-1}(2)} \dots m_n^{\sigma^{-1}(n)} \\
&= \sum_{\sigma} \operatorname{sgn}(\sigma) m_1^{\sigma(1)} m_2^{\sigma(2)} \dots m_n^{\sigma(n)} \\
&= \det M^T.
\end{aligned}$$

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:

$$\operatorname{adj} M = (\operatorname{cofactor}(m^i_j))^T$$

Example

$$\operatorname{adj} \begin{pmatrix} 3 & -1 & -1 \\ 1 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} \det \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} & -\det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \det \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \\ -\det \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix} & \det \begin{pmatrix} 3 & -1 \\ 0 & 1 \end{pmatrix} & -\det \begin{pmatrix} 3 & -1 \\ 0 & 1 \end{pmatrix} \\ \det \begin{pmatrix} -1 & -1 \\ 2 & 0 \end{pmatrix} & -\det \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix} & \det \begin{pmatrix} 3 & -1 \\ 1 & 2 \end{pmatrix} \end{pmatrix}^T$$

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:

$$\begin{aligned} \begin{pmatrix} 3 & -1 & -1 \\ 1 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 2 \\ -1 & 3 & -1 \\ 1 & -3 & 7 \end{pmatrix} &= \begin{pmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix} \\ \Rightarrow \begin{pmatrix} 3 & -1 & -1 \\ 1 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix}^{-1} &= \frac{1}{6} \begin{pmatrix} 2 & 0 & 2 \\ -1 & 3 & -1 \\ 1 & -3 & 7 \end{pmatrix} \end{aligned}$$

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.

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:

$$\text{Volume} = |\det \begin{pmatrix} u & v & w \end{pmatrix}|$$

References

Hefferon, Chapter Four, Section I.1 and I.3

Wikipedia:

- Determinant
- Elementary Matrix
- Cramer's Rule

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.

i. How many additions and multiplications does it take to compute the determinant of a general 2×2 matrix?

ii. 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.

iii. 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 Eigenvalues and Eigenvectors

Matrix of a Linear Transformation Consider a linear transformation

$$L : \mathbb{R}^2 \to \mathbb{R}^2 .$$

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"* $\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\}$.

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 $\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\}$. 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.

15.1 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!*

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.

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.

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 $L:\mathbb{R}^2\to\mathbb{R}^2$ such that $L(x,y)=(2x+2y,16x+6y)$. First, we can find the matrix of L:

$$\begin{pmatrix}x\\y\end{pmatrix}\overset{L}{\longmapsto}\begin{pmatrix}2&2\\16&6\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}.$$

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

$$\begin{pmatrix}4 & 2\\ 16 & 4\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}.$$

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[4] $\det(\lambda I - M)$.
- 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

Wikipedia:

- Eigen*
- Characteristic Polynomial
- Linear Transformations (and matrices thereof)

Review Questions

1. Consider $L : \mathbb{R}^2 \to \mathbb{R}^2$ with $L(x, y) = (x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta)$.

   i. Write the matrix of L on the basis $\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}$.

   ii. When $\theta \neq 0$, explain how L acts on the plane. Draw a picture.

   iii. Do you expect L to have invariant directions?

   iv. Try to find eigenvalues for L by solving the equation

$$L(v) = \lambda v.$$

[4] It is often easier (and equivalent if you only need the roots) to compute $\det(M - \lambda I)$.

v. Does L have real eigenvalues? If not, are there complex eigenvalues for L, assuming that $i = \sqrt{-1}$ exists?

2. Let $M = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$. Find all eigenvalues of M. Does M have two independent eigenvectors? Can M be diagonalized?

3. Let L be the linear transformation $L : \mathbb{R}^3 \to \mathbb{R}^3$ given by $L(x, y, z) = (x + y, y + z, x + z)$. Let e_i be the vector with a one in the ith position and zeros in all other positions.

   i. Find Le_i for each i.

   ii. 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?

   iii. Find a 3×3 matrix M representing L. Choose three non-trivial vectors pointing in different directions and show that $Mv = Lv$ for each of your choices.

16 Eigenvalues and Eigenvectors II

Last time, 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 $L : V \to V$, 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 $L : V \to V$.

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.*

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Then to find the matrix of any linear transformation $L : \mathbb{R}^n \to \mathbb{R}^n$, 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

$$P_M(\lambda) = \lambda^n + c_1\lambda^{n-1} + \ldots + c_n.$$

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

$$P_M(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_n), \qquad P_M(\lambda_i) = 0$$

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 $L : \mathbb{R}^3 \to \mathbb{R}^3$ 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:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \overset{L}{\mapsto} \begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Then the characteristic polynomial of L is[5]

$$\begin{aligned} P_M(\lambda) &= \det \begin{pmatrix} \lambda - 2 & -1 & 1 \\ -1 & \lambda - 2 & 1 \\ 1 & 1 & \lambda - 2 \end{pmatrix} \\ &= (\lambda - 2)[(\lambda - 2)^2 - 1] + [-(\lambda - 2) - 1] + [-(\lambda - 2) - 1] \\ &= (\lambda - 1)^2(\lambda - 4) \end{aligned}$$

[5] 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:

$$\{s \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} |s, t \in \mathbb{R}\}.$$

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 $f : \mathbb{R} \to \mathbb{R}$. Then the derivative is a linear operator $\frac{\partial}{\partial x} : V \to V$. 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 $\frac{\partial}{\partial x}$ if there exists some number λ such that $\frac{\partial}{\partial x} f = \lambda f$. An obvious candidate is the exponential function, $e^{\lambda x}$; indeed, $\frac{\partial}{\partial x} e^{\lambda x} = \lambda e^{\lambda x}$.

As such, the operator $\frac{\partial}{\partial x}$ has an eigenvector $e^{\lambda x}$ for every $\lambda \neq 0 \in \mathbb{R}$. For $\lambda = 0$, $\frac{\partial}{\partial x}$ still has an eigenvector: the zero-function.

This is actually the whole collection of eigenvectors for $\frac{\partial}{\partial x}$; 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 $\frac{\partial}{\partial x}$ with eigenvalues λ, then they are scalar multiples of each other.

16.1 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 $c_1 v_1 + c_2 v_2 + \ldots$ for some constants $\{c_1, c_2, \ldots\}$. Then:

$$\begin{aligned}
L(c_1 v_1 + c_2 v_2 + \ldots) &= c_1 L v_1 + c_2 L v_2 + \ldots \text{ by linearity of } L \\
&= c_1 \lambda v_1 + c_2 \lambda v_2 + \ldots \text{ since } L v_i = \lambda v_i \\
&= \lambda(c_1 v_1 + c_2 v_2 + \ldots).
\end{aligned}$$

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.

References

- Hefferon, Chapter Three, Section III.1: Representing Linear Maps with Matrices
- Hefferon, Chapter Five, Section II.3: Eigenvalues and Eigenvectors

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? Investigate whether something similar holds for $n \times n$ matrices.

17 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!

17.1 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 17.1 (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. The proof is left as an exercise to the reader. □

Note that the requirements of the subspace theorem are often referred to as "closure".

17.2 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:

$$\operatorname{span}(U) = \{x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} | x, y \in \mathbb{R}\}.$$

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:

$$\operatorname{span}(U) = \{r^1 s_1 + r^2 s_2 + \ldots + r^N s_N | r^i \in \mathbb{R}, N \in \mathbb{N}\}.$$

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 span(S) are linear combinations of vectors in the x-axis and the vector P.

Since the sum of any number of vectors along the x-axis is still a vector along the x-axis, then the elements of S are all of the form:

$$\begin{pmatrix}x\\0\\0\end{pmatrix} + y\begin{pmatrix}0\\1\\0\end{pmatrix} = \begin{pmatrix}x\\y\\0\end{pmatrix}.$$

Then span(S) is the xy-plane, which is a vector space. (Try drawing a picture to verify this!)

Lemma 17.2. *For any subset $S \subset V$,* span(S) *is a subspace of V.*

Proof. We need to show that span(S) is a vector space.

It suffices to show that span(S) is closed under linear combinations. Let $u, v \in \operatorname{span}(S)$ and λ, μ be constants. By the definition of span(S), there are constants c^i and d^i (some of which could be zero) such that:

$$\begin{array}{rcl} u & = & c^1 s_1 + c^2 s_2 + \ldots \\ v & = & d^1 s_1 + d^2 s_2 + \ldots \\ \Rightarrow \lambda u + \mu v & = & \lambda(c^1 s_1 + c^2 s_2 + \ldots) + \mu(d^1 s_1 + d^2 s_2 + \ldots) \\ & = & (\lambda c^1 + \mu d^1) s_1 + (\lambda c^2 + \mu d^2) s_2 + \ldots \end{array}$$

This last sum is a linear combination of elements of S, and is thus in span(S). Then span(S) is closed under linear combinations, and is thus a subspace of V. □

Note that this proof, like many proofs, consisted of little more than just writing out the definitions.

Example For which values of a does

$$\operatorname{span}\{\begin{pmatrix}1\\0\\a\end{pmatrix}, \begin{pmatrix}1\\2\\-3\end{pmatrix}, \begin{pmatrix}a\\1\\0\end{pmatrix}\} = \mathbb{R}^3?$$

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

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 $P_3[x]$ be the vector space of degree 3 polynomials 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:

   i. $U \cup W$

   ii. $U \cap W$

   also subspaces? Explain why or why not. Draw examples in $\mathbb{R}^3$.

18 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 certainly $P = \operatorname{span}\{u, v, w\}$. 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 $c^1, c^2, \ldots, c^n$ not all zero such that

$$c^1v_1 + c^2v_2 + \ldots + c^nv_n = 0.$$

Otherwise, the vectors $v_1, v_2, \ldots, v_n$ are *linearly independent.*

Example Consider the following vectors in $\mathbb{R}^3$:

$$v_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Are they linearly independent?

We need to see whether the system

$$c^1v_1 + c^2v_2 + c^3v_3 = 0.$$

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:

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

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 independence 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 $c^3 = -2\mu$. 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:

$$c^1 v_1 + c^2 v_2 + c^3 v_3 = 0 \Rightarrow -2v_1 - v_2 + v_3 = 0.$$

Theorem 18.1 (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 $v_k = c^1 v_1 + \ldots c^{k-1} v_{k-1}$ then the set is linearly dependent.

This is easy. We just rewrite the assumption:

$$c^1 v_1 + \ldots c^{k-1} v_{k-1} - v_k + 0 v_{k+1} + \ldots + 0 v_n = 0.$$

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

$$c^1v_1 + c^2v_2 + \ldots + c^nv_n = 0.$$

Take k to be the largest number for which c_k is not equal to zero. So:

$$c^1v_1 + c^2v_2 + \ldots + c^{k-1}v_{k-1} + c^kv_k = 0.$$

(Note that $k > 1$, since otherwise we would have $c^1v_1 = 0 \Rightarrow v_1 = 0$, contradicting the assumption that none of the v_i are the zero vector.)

As such, we can rearrange the equation:

$$\begin{array}{rcl} c^1v_1 + c^2v_2 + \ldots + c^{k-1}v_{k-1} & = & -c^kv_k \\ \Rightarrow \ -\frac{c^1}{c^k}v_1 - \frac{c^2}{c^k}v_2 - \ldots - \frac{c^{k-1}}{c^k}v_{k-1} & = & v_k. \end{array}$$

Therefore we have expressed v_k as a linear combination of the previous vectors, and we are done.

□

Example Consider the vector space $P_2(t)$ of polynomials of degree less than or equal to 2. Set:

$$\begin{array}{rcl} v_1 & = & 1+t \\ v_2 & = & 1+t^2 \\ v_3 & = & t+t^2 \\ v_4 & = & 2+t+t^2 \\ v_5 & = & 1+t+t^2. \end{array}$$

The set $\{v_1, \ldots, v_5\}$ is linearly dependent, because $v_4 = v_1 + v_2$.

Now suppose vectors $v_1, \ldots, v_n$ are linearly dependent,

$$c^1v_1 + c^2v_2 + \ldots + c^nv_n = 0$$

with $c^1 \neq 0$. Then:

$$\mathrm{span}\{v_1, \ldots, v_n\} = \mathrm{span}\{v_2, \ldots, v_n\}$$

because any $x \in \operatorname{span}\{v_1, \ldots, v_n\}$ is given by

$$\begin{aligned}
x &= a^1 v_1 + \ldots a^n v_n \\
&= a^1\left(-\frac{c^2}{c_1} v_2 - \ldots - \frac{c^n}{c_1} v_n\right) + a^2 v_2 + \ldots + a^n v_n \\
&= \left(a^2 - a^1 \frac{c^2}{c_1}\right) v_2 + \ldots + \left(a^n - a^1 \frac{c^n}{c_1}\right) v_n .
\end{aligned}$$

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. 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

$$\begin{pmatrix}1 & 1 & 1\\1 & 0 & 1\\0 & 1 & 1\end{pmatrix} \begin{pmatrix}c^1\\c^2\\c^3\end{pmatrix} = 0.$$

Solutions exist if and only if the determinant of the matrix is non-zero. But:

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

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

Wikipedia:

- Linear Independence
- Basis

Review Questions

1. Let B^n be the space of $n \times 1$ bit-valued matrices (*i.e.*, column vectors).

   i. How many different vectors are there in B^n.

   ii. Find a collection S of vectors that span B^3 and are linearly independent. In other words, find a basis of B^3.

iii. Write each other vector in B^3 as a linear combination of the vectors in the set S that you chose.

iv. 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 $\mathbb{R}^n$.

    i. Prove that the collection $\{e_1, \ldots, e_n\}$ is linearly independent.

    ii. Show that $v = \sum_{i=1}^{n}(v \cdot e_i)e_i$.

    iii. What does this say about the span$\{e_1, \ldots, e_n\}$?

19 Basis and Dimension

In the last section, 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 $\operatorname{span} S = V$.

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

$$a^0\,1 + a^1\,t + \ldots + a^n\,t^n, \qquad a^i \in \mathbb{R}$$

so $P_n(t) = \operatorname{span}\{1, t, \ldots, t^n\}$. This set of vectors is linearly independent: If the polynomial $p(t) = c^0 1 + c^1 t + \ldots + c^n t^n = 0$, then $c^0 = c^1 = \ldots = c^n = 0$, so $p(t)$ is the zero polynomial.

Then $P_n(t)$ is finite dimensional, and $\dim P_n(t) = n + 1$.

Theorem 19.1. *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:*

$$w = c^1 v_1 + \ldots + c^n v_n.$$

Proof. Since S is a basis for V, then $\operatorname{span} S = V$, and so there exists constants c^i such that $w = c^1 v_1 + \ldots + c^n v_n$.

Suppose there exists a second set of constants d^i such that

$$w = d^1 v_1 + \ldots + d^n v_n \,.$$

Then:

$$\begin{array}{rcl} 0_V & = & w - w \\ & = & c^1 v_1 + \ldots + c^n v_n - d^1 v_1 + \ldots + d^n v_n \\ & = & (c^1 - d^1) v_1 + \ldots + (c^n - d^n) v_n . \end{array}$$

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, $c^i = d^i$. □

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 19.2. *If $S = \{v_1, \ldots, v_n\}$ is a basis for a vector space V and $T = \{w_1, \ldots, w_m\}$ is a linearly independent set of vectors in V, then $m \leq n$.*

Proof. The idea 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.

Since S spans V, then the set $\{w_1, v_1, \ldots, v_n\}$ 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

$$v_k = c^0 w_1 + c^1 v_1 + \ldots + c^{i-1} v_{i-1} + c^{i+1} v_{i+1} + \ldots + c^n v_n.$$

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 19.3. *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. □

19.1 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

$$0 = x^1 v_1 + \ldots + x^n v_n.$$

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

$$w = x^1 v_1 + \ldots + x^n v_n$$

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 19.4. *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

$$S = \{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\} \text{ and } T = \{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}\}.$$

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

Wikipedia:

- Linear Independence
- Basis

Review Questions

1. Let S be a collection of vectors 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. (This is the converse to the theorem in the lecture.)

2. 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 $T : \mathbb{R}^3 \to \mathbb{R}$ is just a column vector.)

   (Hint: If you are really stuck (or just curious), look up "dual space." This is a big idea, though, and could just be more confusing.)

3. *i.* Draw the collection of all unit vectors in $\mathbb{R}^2$.

   ii. Let $S_x = \{\begin{pmatrix}1\\0\end{pmatrix}, x\}$, where x is a unit vector in $\mathbb{R}^2$. For which x is S_x a basis of $\mathbb{R}^2$?

4. 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 $k \leq n$?)

20 Diagonalization

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:

$$\begin{aligned} v &= v^1 e_1 \ + v^2 e_2 \ + \ldots + v^n e_n \\ w &= w^1 f_1 + w^2 f_2 + \ldots + w^m f_m \end{aligned}$$

We call the coefficients $v^1, \ldots, v^n$ the *components* of v in the basis[6] $\{e_1, \ldots, e_n\}$. 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 $\mathbb{R}^m$:

$$\begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \in \mathbb{R}^n, \qquad \begin{pmatrix} w^1 \\ \vdots \\ w^m \end{pmatrix} \in \mathbb{R}^m.$$

Now suppose we have a linear transformation $L : V \to W$. 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.

[6] 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.

Using linearity, we write:

$$\begin{aligned}
L(v) &= L(v^1 e_1 + v^2 e_2 + \ldots + v^n e_n) \\
&= v^1 L(e_1) + v^2 L(e_2) + \ldots + v^n L(e_n) \\
&= \begin{pmatrix} L(e_1) & L(e_2) & \cdots & L(e_n) \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix} .
\end{aligned}$$

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_j^i such that:

$$L(e_j) = f_1 M_j^1 + \ldots + f_m M_j^m = \sum_{i=1}^{m} f_i M_j^i = \begin{pmatrix} f_1 & f_2 & \cdots & f_m \end{pmatrix} \begin{pmatrix} M_j^1 \\ M_j^2 \\ \vdots \\ M_j^m \end{pmatrix} .$$

We've written the M_j^i 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 left, like so: $f_i M_j^i$.

Now M_j^i is the ith component of $L(e_j)$. Regarding the coefficients M_j^i 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:

$$\begin{aligned}
L(v) &= L(v^1 e_1 + v^2 e_2 + \ldots + v^n e_n) \\
&= v^1 L(e_1) + v^2 L(e_2) + \ldots + v^n L(e_n) \\
&= \sum_{i=1}^{m} v^j L(e_j) \\
&= \sum_{i=1}^{m} v^j (M_j^1 f_1 + \ldots + M_j^m f_m) \\
&= \sum_{i=1}^{m} f_i [\sum_{j=1}^{n} M_j^i v^j] \\
&= \begin{pmatrix} f_1 & f_2 & \cdots & f_m \end{pmatrix} \begin{pmatrix} M_1^1 & M_2^1 & \cdots & M_n^1 \\ M_1^2 & M_2^2 & & \\ \vdots & & & \vdots \\ M_1^m & & \cdots & M_n^m \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix}
\end{aligned}$$

The second last equality is the definition of matrix multiplication which is obvious from the last line. Thus:

$$\begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \overset{L}{\mapsto} \begin{pmatrix} M_1^1 & \ldots & M_n^1 \\ \vdots & & \vdots \\ M_1^m & \ldots & M_n^m \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix},$$

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 $L : P_1(t) \mapsto P_1(t)$, 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.

Now suppose we are lucky, and we have $L : V \mapsto V$, 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.

We call the $n \times n$ matrix of a linear transformation $L : V \mapsto V$ *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 $L : V \to V$, then the matrix for L in the basis S is diagonal if and only if S is a basis of eigenvectors for L.*

20.1 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. To wit, suppose $L : V \mapsto V$ 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

$$L(v_i) = \lambda_i v_i = \sum_k v_k D^k_i$$

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:

$$L(v_j) = L(\sum_i u_i P^i_j) = \sum_i L(u_i) P^i_j = \sum_i \sum_k u_k M^k_i P^i_j.$$

Meanwhile, we have:

$$L(v_i) = \sum_k v_k D^k_i = \sum_k \sum_j u_j P^j_k D^k_i.$$

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.*

References

- Hefferon, Chapter Three, Section V: Change of Basis

Wikipedia:

- Change of Basis
- Diagonalizable Matrix
- Similar Matrix

Review Questions

1. Show that similarity of matrices is an *equivalence relation*. (The definition of an equivalence relation is given in Section 2, in the fourth review problem.)

2. When is the 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ diagonalizable? Include examples in your answer.

3. Let $P_n(t)$ be the vector space of degree n polynomials, and $\frac{d}{dt} : P_n(t) \mapsto P_{n-1}(t)$ 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 $P_{n-1}(t)$.

4. 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. (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. Can you make any observations about the resulting matrices?

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

$$\begin{aligned}
\Pi_1 &= e_1 e_1^T \\
&= \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \\
&\vdots \\
\Pi_n &= e_n e_n^T \\
&= \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & \dots & 1 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}
\end{aligned}$$

In short, Π_i is the diagonal square matrix with a 1 in the ith diagonal position and zeros everywhere else. [7]

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 $\lambda_1, \dots, \lambda_n$, we can write

$$D = \lambda_1 \Pi_1 + \dots + \lambda_n \Pi_n.$$

[7] 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:

$$v = c^1 u_1 + \ldots c^n u_n.$$

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:

$$\begin{array}{rcl} v \cdot u_i &=& c^1 u_1 \cdot u_i + \ldots + c^i u_i \cdot u_i + \ldots + c^n u_n \cdot u_i \\ &=& c^1 \cdot 0 + \ldots + c^i \cdot 1 + \ldots + c^n \cdot 0 \\ &=& c^i, \\ \Rightarrow c^i &=& v \cdot u_i \\ \Rightarrow v &=& (v \cdot u_1) u_1 + \ldots + (v \cdot u_n) u_n \\ &=& \sum_i (v \cdot u_i) u_i. \end{array}$$

This proves the theorem:

Theorem 21.1. *For an orthonormal basis $\{u_1, \ldots, u_n\}$, any vector v can be expressed*

$$v = \sum_i (v \cdot u_i) u_i.$$

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:

$$\begin{array}{rcl} w_1 & = & (w_1 \cdot u_1)u_1 + \ldots + (w_1 \cdot u_n)u_n \\ & \vdots & \\ w_n & = & (w_n \cdot u_1)u_1 + \ldots + (w_n \cdot u_n)u_n \\ \Rightarrow w_i & = & \sum_j u_j(u_j \cdot w_i) \end{array}$$

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.

$$\begin{array}{rcl} (PP^T)_k^j & = & \sum_i (u_j \cdot w_i)(w_i \cdot u_k) \\ & = & \sum_i (u_j^T w_i)(w_i^T u_k) \\ & = & u_j^T \left[\sum_i (w_i w_i^T) \right] u_k \\ & = & u_j^T I_n u_k \qquad (*) \\ & = & u_j^T u_k = \delta_{jk}. \end{array}$$

In 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 $(\sum_i w_i w_i^T)v$ for an arbitrary vector v. We can find constants c^j such that $v = \sum_j c^j w_j$, so that:

$$\begin{aligned}
(\sum_i w_i w_i^T)v &= (\sum_i w_i w_i^T)(\sum_j c^j w_j) \\
&= \sum_j c^j \sum_i w_i w_i^T w_j \\
&= \sum_j c^j \sum_i w_i \delta_{ij} \\
&= \sum_j c^j w_j \text{ since all terms with } i \neq j \text{ vanish} \\
&= v.
\end{aligned}$$

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.*

$$P^{-1} = P^T.$$

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 $P^TP = I$ by a lengthy computation, or more simply, notice that

$$\begin{aligned} (P^TP)_{ij} &= \begin{pmatrix} u_1^T \\ u_2^T \\ u_3^T \end{pmatrix} \begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

We are using orthonormality of the u_i for the matrix multiplication above.

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

Wikipedia:

- Orthogonal Matrix
- Diagonalizable Matrix
- Similar Matrix

Review Questions

1. Let $D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$.

   i. Write D in terms of the vectors e_1 and e_2, and their transposes.

   ii. Suppose $P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible. Show that D is similar to

   $$M = \frac{1}{ad - bc} \begin{pmatrix} \lambda_1 ad - \lambda_2 bc & (\lambda_1 - \lambda_2)bd \\ (\lambda_1 - \lambda_2)ac & -\lambda_1 bc + \lambda_2 ad \end{pmatrix}.$$

   iii. 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?

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.

   i. Is the vector $v^\perp = v - \frac{u \cdot v}{u \cdot u} u$ in the plane P?

   ii. What is the angle between $v^\perp$ and u?

   iii. Given your solution to the above, how can you find a third vector perpendicular to both u and $v^\perp$?

iv. Construct an orthonormal basis for $\mathbb{R}^3$ from u and v.

v. Test your abstract formulae starting with

$$u = \begin{pmatrix} 1 & 2 & 0 \end{pmatrix} \text{ and } v = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}.$$

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:

$$Mx = \lambda x, \qquad My = \lambda y.$$

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 21.3. *Eigenvectors of a symmetric matrix with distinct eigenvalues are orthogonal.*

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:

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0$$

Last lecture we saw that the matrix P built from orthonormal basis vectors $\{v_1, \ldots, v_n\}$

$$P = \begin{pmatrix} v_1 & \ldots & v_n \end{pmatrix}$$

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, $\dots$, x_n such that $\{x_1, \dots, x_n\}$ is an orthonormal basis. (Such a basis can be obtained using the "Gram-Schmidt" procedure, which we will present later.)

Hyperlink to notes24!

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:

$$P = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix},$$

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

$$MP = \begin{pmatrix} \lambda_1 x_1 & Mx_2 & \dots & Mx_n \end{pmatrix}.$$

But P is an orthogonal matrix, so $P^{-1} = P^T$. Then:

$$\begin{aligned}
P^{-1} = P^T &= \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} \\
\Rightarrow P^T M P &= \begin{pmatrix} x_1^T \lambda_1 x_1 & * & \dots & * \\ x_2^T \lambda_1 x_1 & * & \dots & * \\ \vdots & & & \vdots \\ x_n^T \lambda_1 x_1 & * & \dots & * \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 & * & \dots & * \\ 0 & * & \dots & * \\ \vdots & * & & \vdots \\ 0 & * & \dots & * \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \hat{M} & \\ 0 & & & \end{pmatrix}
\end{aligned}$$

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 21.4. *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.

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

Wikipedia:

- Symmetric Matrix
- Diagonalizable Matrix
- Similar Matrix

Review Questions

1. (On Reality of Eigenvectors)

   i. Suppose $z = x + iy$ where $x, y \in \mathbb{R}, i = \sqrt{-1}$, and $\overline{z} = x - iy$. Compute $\overline{\overline{z}}$. What can you say about $z\overline{z}$ and $\overline{z}z$? This operation is called *complex conjugation*.

   ii. What can you say about complex numbers λ that obey $\lambda = \overline{\lambda}$?

   iii. Let $x = \begin{pmatrix} z^1 \\ \vdots \\ z^n \end{pmatrix} \in \mathbb{C}^n$. Let $x^\dagger = \begin{pmatrix} \overline{z^1} & \dots & \overline{z^n} \end{pmatrix} \in \mathbb{C}^n$. Compute $x^\dagger x$. What can you say about the result?

   iv. 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:

   $$\frac{x^\dagger M x}{x^\dagger x}$$

   v. Suppose Λ is a 1×1 matrix. What is Λ^T?

   vi. What is the size of the matrix $x^\dagger M x$?

   vii. 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 $\overline{(x^\dagger M x)^T}$.

   viii. Show that $\lambda = \overline{\lambda}$. 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. What can you say about the sum of the dimensions of the eigenspaces of a real symmetric matrix?

22 Kernel, Range, Nullity, Rank

Define all for functions; segue into linear transformations; merge footnote into text. Fluffy text before hitting with definitions.

The *range* of a linear transformation $L : V \to W$ is the set of vectors the linear transformation maps to. This set is also often called the *image* of L, written

$$\operatorname{ran}(L) = \operatorname{Im}(L) = L(V) = \{L(v) | v \in V\} \subset W.$$

The *domain* V of a linear transformation $L : V \to W$ is often called the *pre-image* of L. We can also talk about the pre-image of any subset of vectors $U \in W$:

$$L^{-1}(U) = \{v \in V | L(v) \in U\} \subset V.$$

A linear transformation L is *one-to-one* if for any $x \neq y \in V$, $L(x) \neq f(y)$. In other words, different vectors in V always map to different vectors in W. One-to-one transformations are also known as *injective* transformations. Notice that injectivity is a condition on the pre-image of L.

A linear transformation L is *onto* if for every $w \in W$, there exists an $x \in V$ such that $L(x) = w$. In other words, every vector in W is the image of some vector in V. An onto transformation is also known as an *surjective* transformation. Notice that surjectivity is a condition on the image of L. [8]

Suppose $L : V \to W$ 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 $L : V \to W$ be a linear transformation. The set of all vectors v such that $Lv = 0_W$ is called the *kernel of* L:

$$\ker L = \{v \in V | Lv = 0_W\}.$$

[8] The notions of one-to-one and onto can be generalized to arbitrary functions on sets. For example if g is a function from a set S to a set T, then g is one-to-one if different objects in S always map to different objects in T. For a linear transformation f, these sets S and T are then just vector spaces, and we require that f is a linear map; *i.e.* f respects the linear structure of the vector spaces.

The linear structure of sets of vectors lets us say much more about one-to-one and onto functions than one can say about functions on general sets. For example, we always know that a linear function sends 0_V to 0_W. Then we can 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!

Theorem 22.1. *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.

Theorem 22.2. *Let $L : V \to W$. 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. □

hyperlink to subspace theorem

This theorem has an interpretation in terms of the eigenspaces of $L : V \to V$. 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 $\mathbb{R}^3$.

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 22.3. *Let $L : V \to W$. 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

$$c^1 L(v_1) + \ldots + c^n L(v_n) = 0.$$

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 22.4 (Dimension Formula)**.** *Let $L : V \to W$ be a linear transformation, with V a finite-dimensional vector space*[9]*. 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 $L(V)$.

[9] 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.

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:

$$\begin{array}{rcl} w & = & L(c^1 v_1 + \ldots + c^p v_p + d^1 u_1 + \ldots + d^q u_q) \\ & = & c^1 L(v_1) + \ldots + c^p L(v_p) + d^1 L(u_1) + \ldots + d^q L(u_q) \\ & = & d^1 L(u_1) + \ldots + d^q L(u_q) \text{ since } L(v_i) = 0, \\ \Rightarrow L(V) & = & \operatorname{span}\{L(u_1), \ldots, L(u_q)\}. \end{array}$$

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

$$\begin{array}{rcl} 0 & = & d^1 L(u_1) + \ldots + d^q L(u_q) \\ & = & L(d^1 u_1 + \ldots + d^q u_q). \end{array}$$

But since the u^j are linearly independent, then $d^1 u_1 + \ldots + d^q u_q \neq 0$, and so $d^1 u_1 + \ldots + d^q u_q$ is in the kernel of L. But then $d^1 u_1 + \ldots + d^q u_q$ 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. □

References

- Hefferon, Chapter Three, Section II.2: Rangespace and Nullspace (Recall that "homomorphism" is is used instead of "linear transformation" in Hefferon.)

Wikipedia:

- Rank
- Dimension Theorem
- Kernel of a Linear Operator

Review Questions

1. Let $L : V \to W$ be a linear transformation. Prove that $\ker L = \{0_V\}$ if and only if L is one-to-one.

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 $L : \mathbb{R}^4 \to \mathbb{R}^3$ whose matrix M in the standard basis is row equivalent to the folowing 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 $L(V)$.

   Find and describe and algorithm (*i.e.* a general procedure) for finding a basis for $L(V)$ when $L : \mathbb{R}^n \to \mathbb{R}^m$.

   Finally, provide an example of the use of your algorithm.

   Hold hands on this a bit more!

4. Claim: If $\{v_1, \ldots, v_n\}$ is a basis for $\ker L$, where $L : V \to W$, 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.

23 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:

$$v^{\perp} = v - \frac{u \cdot v}{u \cdot u} u .$$

Put pictures here?

This new vector $v^{\perp}$ is orthogonal to u because

$$u \bullet v^{\perp} = u \bullet v - \frac{u \cdot v}{u \cdot u} u \bullet u = 0 .$$

Hence, $\{u, v^{\perp}\}$ is an orthogonal basis for span$\{u, v\}$. When v is not parallel to u, $v^{\perp} \neq 0$, and normalizing these vectors we obtain $\{\frac{u}{|u|}, \frac{v^{\perp}}{|v^{\perp}|}\}$, 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. [10]

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:

$$w^{\perp} = w - \frac{u \cdot w}{u \cdot u}\, u - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}}\, v^{\perp} .$$

[10] 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.

One can check by directly computing $u \cdot w^{\perp}$ and $v^{\perp} \cdot w^{\perp}$ that $w^{\perp}$ is orthogonal to both u and $v^{\perp}$; as such, $\{u, v^{\perp}, w^{\perp}\}$ is an orthogonal basis for $\operatorname{span}\{u, v, w\}$.

Put in the computation!

In fact, given a collection $\{v_1, v_2, \ldots\}$ of linearly independent vectors, we can produce an orthogonal basis for $\operatorname{span}\{v_1, v_2, \ldots\}$ consisting of the following vectors:

$$\begin{array}{rcl}
v_1^{\perp} & = & v_1 \\
v_2^{\perp} & = & v_2 - \dfrac{v_1 \cdot v_2}{v_1 \cdot v_1}\, v_1 \\
 & \vdots & \\
v_3^{\perp} & = & v_3 - \dfrac{v_1^{\perp} \cdot v_3}{v_1^{\perp} \cdot v_1^{\perp}}\, v_1^{\perp} - \dfrac{v_2^{\perp} \cdot v_3}{v_2^{\perp} \cdot v_2^{\perp}}\, v_2^{\perp} \\
 & \vdots & \\
v_i^{\perp} & = & v_i - \displaystyle\sum_{j<i} \dfrac{v_j^{\perp} \cdot v_i}{v_j^{\perp} \cdot v_j^{\perp}}\, v_j^{\perp} \\
 & = & v_i - \dfrac{v_1^{\perp} \cdot v_i}{v_1^{\perp} \cdot v_1^{\perp}}\, v_1^{\perp} - \ldots - \dfrac{v_{n-1}^{\perp} \cdot v_i}{v_{n-1}^{\perp} \cdot v_{n-1}^{\perp}}\, v_{n-1}^{\perp} \\
 & \vdots &
\end{array}$$

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:

$$\begin{array}{rcl}
v^{\perp} & = & \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} - \frac{2}{2}\begin{pmatrix} 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \\
w^{\perp} & = & \begin{pmatrix} 3 & 1 & 1 \end{pmatrix} - \frac{4}{2}\begin{pmatrix} 1 & 1 & 0 \end{pmatrix} - \frac{1}{1}\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}.
\end{array}$$

Then the set

$$\{\begin{pmatrix} 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}\}$$

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:

$$\left\{\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\end{pmatrix}, \begin{pmatrix}0 & 0 & 1\end{pmatrix}, \begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0\end{pmatrix}\right\}.$$

23.1 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

$$\operatorname{span} U \cup V = \{u + v | u \in U, v \in V\} = U + V$$

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:

$$U \oplus V = \operatorname{span} U \cup V = \{u + v | u \in U, v \in V\}.$$

The direct sum has a very nice property.

Theorem 23.1. *Let $w = u + v \in U \oplus V$. Then the expression $w = u + v$ is unique.*

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. □

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.

Theorem 23.2. *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:

$$\begin{array}{rcl} u & = & (w \cdot e_1)e_1 + \ldots + (w \cdot e_n)e_n \in U \\ u^{\perp} & = & w - u \end{array}$$

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. □

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

$$\mathbb{R}^3 = P \oplus P^{\perp} = \{(x,y,0)|x,y \in \mathbb{R}\} \oplus \{(0,0,z)|z \in \mathbb{R}\}.$$

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

Wikipedia:

- Gram-Schmidt Process
- Orthonormal Basis
- Direct Sum

Review Questions

1. 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 span$\{u, v\}$.

2. Repeat the previous problem, but with three independent vectors u, v, w, and $v^{\perp}$ and $w^{\perp}$ as defined in the lecture.

3. Given any three vectors u, v, w, when do $v^{\perp}$ or $w^{\perp}$ vanish?

4. This question will answer the question, "If I choose a vector *at random*, what is the probability that it lies in the span of some other vectors?"

5. For U a subspace of W, use the subspace theorem to check that $U^{\perp}$ is a subspace of W.

   i. Given a collection S of k vectors in $\mathbb{R}^n$, consider the 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 vector v. Suppose S is a collection of k linearly independent vectors in $\mathbb{R}^n$. 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. Let M be an $n \times n$ diagonalizable matrix whose eigenvalues are chosen uniformly at random. (*i.e.* every real number has equal chance of being an eigenvalue.) What is the probability that the columns of M form a basis for $\mathbb{R}^n$? (Hint: What is the relationship between the kernel of M and its eigenvalues?)

24 Least Squares

Consider the linear system $L(x) = v$, where $L : U \xrightarrow{\text{linear}} W$, 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, …, Me_n.) Then $v^\perp$ must be perpendicular to the columns of M. *i.e.*, $M^T(V - MX) = 0$, or

$$M^TMX = M^TV.$$

Solutions X to $M^TMX = M^TV$ 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 $M^TMX = M^TV$.

Observe that since M is an $m \times n$ matrix, then M^T is an $n \times m$ matrix. Then M^TM is an $n \times n$ matrix, and is symmetric, since $(M^TM)^T = M^TM$. Then, for any vector X, we can evaluate X^TM^TMX to obtain a number. This is a very nice number, though! It is just the length $|MX|^2 = (MX)^T(MX) = X^TM^TMX$.

Now suppose that $\ker L = \{0\}$, so that the only solution to $MX = 0$ is $X = 0$; in particular, M is invertible. But if M is invertible, then so is M^TM, since $(M^TM)^{-1} = M^{-1}M^{-T}$. Then the only solution to $M^TMX = 0$ is $X = 0$.

In this case, the least squares solution (the X that solves $M^TMX = MV$) is unique, and is equal to

$$X = (M^TM)^{-1}M^TV.$$

In a nutshell, this is the least squares method.

- Compute M^TM and M^TV.
- Solve $(M^TM)X = M^TV$ 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[11], he believes that his data are best approximated by a straight line

$$v = at + b.$$

Then he should find a and b to best fit the data.

$$\begin{aligned} 11 &= a \cdot 1 + b \\ 19 &= a \cdot 2 + b \\ 31 &= a \cdot 3 + b. \end{aligned}$$

[11] In fact, he is a *Calculus Superhero*.

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 $M^TMX = M^TV$.

$$\begin{pmatrix}1&2&3\\1&1&1\end{pmatrix}\begin{pmatrix}1&1\\2&1\\3&1\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}=\begin{pmatrix}1&2&3\\1&1&1\end{pmatrix}\begin{pmatrix}11\\19\\31\end{pmatrix}.$$

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....)!

References

- Hefferon, Chapter Three, Section VI.2: Gram-Schmidt Orthogonalization

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 $L(u_{\text{ps}}) = v$.

   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 $\mathbb{R}^n$.)

Index