

Math 256: Linear Algebra

Dr. Heather Moon

contributing-Authors:

Dr. Thomas J. Asaki,

Dr. Marie Snipes,

and

Dr. Chris Camfield

April 20, 2016

Contents

1	Systems of Equations	5
1.1	Review from your basic Algebra class	5
1.1.1	Substitution	5
1.1.2	Elimination	5
1.2	Systems of Equations with more variables	6
1.3	Using Matrices to solve systems of equations	8
2	Matrix operations and Matrix Equations	12
2.1	Basic Operations on Matrices	12
2.1.1	Matrix Addition	13
2.1.2	Matrix Multiplication	13
2.2	Matrix Equations and more	14
2.3	Determinants	18
3	Radiography and Tomography in Linear Algebra Lab #1	24
3.1	Grayscale Images	24
4	Vector Spaces	26
4.1	Vectors and Vector Spaces	27
4.2	Subspaces	32
5	Span	35
5.1	Definition of Span	36
5.2	Span as a Vector Space	38
6	Linear Independence/Dependence	41
6.1	Linear Dependence Definitions	41
6.2	Basis	43
6.3	Dimension	44
7	Coordinate Spaces	49
7.1	Coordinates in the Standard Basis	49
7.2	Coordinates in Other Bases	50
8	Transmission Radiography and Tomography	
	A Simplified Overview	54
8.1	What is Radiography?	54
8.1.1	Transmission Radiography	55
8.1.2	Tomography	55
8.2	The Incident X-ray Beam	55
8.3	X-Ray Beam Attenuation	56
8.4	Radiographic Energy Detection	57
8.5	The Radiographic Transformation Operator	58
8.6	Multiple Views and Axial Tomography	58
8.7	Model Summary	60
8.8	Model Assumptions	60
8.9	Additional Resources	62

9	Radiography for Linear Algebra Lab #2	63
9.1	Radiographic Scenarios and Notation	63
9.2	A First Example	64
9.3	Now it's your turn!	64
9.3.1	Part I: Finding the matrix associated with a radiographic transformation, and using the matrix to determine radiographic outputs.	65
9.3.2	Part II: Radiographs of Linear Combinations of Objects	67
10	Transformations	67
10.1	Linear Transformations	68
11	Radiography and Tomography in Linear Algebra Lab #3	71
12	Properties of Transformations	73
12.1	Injections and Surjections	73
12.2	Bijections and Isomorphisms	75
13	Transformation Spaces	79
13.1	The Nullspace	79
13.2	Domain and Range Spaces	81
13.3	Injectivity and Surjectivity Revisited	82
13.4	The Rank-Nullity Theorem	85
14	Matrix Representations	87
15	Matrix Spaces	92
15.1	The Nullspace of a Matrix	93
15.2	The Columns Space of a Matrix	95
15.3	Onto, One-to-One, and what this means about Matrices	99
16	Inverse Transformations	100
16.1	A (Very) Brief Introduction to Tomography	100
16.2	Finding solutions to $Mv = w$	102
16.3	Finding solutions to $Mv = e_k$, an element of the standard basis.	103
16.4	The Inverse Matrix, when $m = n$	105
17	Heat States in a (Vector) State Space	106
17.1	The Heat Equation	107
17.2	The Discrete Heat Equation, in x	108
17.3	The Discrete Heat Equation, in t	110
17.4	Heat State Evolution	111
18	Heat Equation for Linear Algebra Lab 1	112
19	Heat Equation for Linear Algebra Lab 2	113

20 Eigenvalues, Eigenvectors, Eigenspaces, and Diagonalizability	114
20.1 Eigenvectors and Eigenvalues	115
20.2 Eigenbasis	116
20.3 Diagonalizable matrices	119
21 Heat Equation for Linear Algebra Lab #3	124
22 Radiography and Tomography in Linear Algebra	
Lab #4	127
22.1 Task 1	127
22.2 Task 2	128
22.3 Task 3	128
23 Special Matrices	128
23.1 Identity Matrix	128
23.2 Diagonal Matrices	129
23.3 Triangular and Block matrices	129
23.3.1 The Transpose and Symmetric Matrices	131
24 Radiography and Tomography in Linear Algebra Lab #5: Reconstruction	
Without an Inverse	132
24.1 Invertible transformation	132
24.2 Application to a small example	133
24.3 Application to Brain Reconstruction	134
25 Inner Products and Projections	136
25.1 Inner Products	136
25.2 Projections	139

1 Systems of Equations

In this section, we will discuss techniques for solving systems of linear equations. There are things to explore in this section. But mostly, I would like you to keep in mind that solution techniques for systems of linear equations are tools we use for more interesting Linear Algebra concepts.

1.1 Review from your basic Algebra class

In high school Algebra classes, students are usually asked to solve systems of equations such as:

$$\begin{array}{r} x + y = 2 \\ x - y = 6 \end{array} .$$

Let's review what we know about these equations.

1. Each of these equations can be plotted on a cartesian (xy -plane). Plot these two.
2. What are the graphs of these two equations?

The solution to the system of equations is where these two graphs cross. We can use one of two different methods to accurately find this point. Of course, you can look back at graphs and estimate the solution and you would probably be correct because I chose the solution to be an integer point. Let's look at two solution techniques.

1.1.1 Substitution

The substitution method works just as the name suggests. Choose one equation and solve for a variable and substitute into the other equation, like so:

$$x + y = 2 \quad \Rightarrow \quad y = 2 - x \quad \Rightarrow \quad x - (2 - x) = 6.$$

Now we solve for x to get $x = 4$. Then we substitute to get $y = 2 - 4 = -2$. So the solution is $(4, -2)$. Remember, the solution is a point where the two graphs cross.

1.1.2 Elimination

We will focus more on this technique in later subsections. The technique, again as the name suggests, requires that we eliminate a variable in order to solve for the other variable. To eliminate a variable, you can add multiples of equations together. In this case, we just want to add the equations together to eliminate y :

$$\begin{array}{r} x + y = 2 \\ + x - y = 6 \\ \hline 2x \qquad = 8 \end{array} .$$

We then solve for x and back substitute. We get $x = 4$ and $y = -2$ as before. So again, we see that the solution is $(4, -2)$.

1.2 Systems of Equations with more variables

In this section, we will continue with the method of elimination, but this time we will solve systems with 3 or more variables. The method of substitution still works, but becomes messier with more variables so we will avoid using that method. Before we go through an example, let's lay out the rules.

Allowed operations when solving systems of linear equations.

- Multiply both sides of an equation by a nonzero number.
- Add multiples of equations together to change only one of them. This means you can replace one equation by the result you get when you multiply two equations by a number and then add them together.
- Change the order of equations.

Using only these operations, let's look at an example. We will solve the system:

$$\begin{array}{r} x + y + z = 0 \\ 2x - y + 3z = -6 \\ -x + y - 2z = 5 \end{array}$$

If you have taken a Vector Calculus class, you know that the graphs of these equations are planes in \mathbb{R}^3 (3-dimensional space). We are looking for the point (or points) where these three planes cross each other. Like I said above, we will use elimination. There are many choices to be made here. I tend to try to eliminate all x 's except in the first equation and then all y 's except in the first two equations. It makes a kind of triangular shape to my system, but mostly, it leaves only z terms in the third equation. We also want to maintain good bookkeeping so, we will always keep the system together. Finally, I want to note that we will make notations to indicate which of the three allowed operations we used in each step. The notation I use (and you should find a notation you like) uses uppercase E to indicate a newly made equation and a lowercase e to indicate equations from the previous step. So, something like $E_2 = 2e_1 + e_2$ means that I will replace equation 2 with a new equation that is formed by taking twice the first equation added to the second.

Example 1.1. *Let's solve the above system now:*

$$\begin{array}{r} x + y + z = 0 \\ 2x - y + 3z = -6 \\ -x + y - 2z = 5 \end{array} \quad \begin{array}{l} E_2 = -2e_1 + e_2 \\ E_3 = e_1 + e_3 \end{array} \quad \begin{array}{r} x + y + z = 0 \\ -3y + z = -6 \\ 2y - z = 5 \end{array}$$

$$\begin{array}{r} x + y + z = 0 \\ y = 1 \\ 2y - z = 5 \end{array} \quad \begin{array}{l} E_2 = -e_2 - e_3 \end{array}$$

It turns out that in my attempt to follow the above strategy, I found y sooner than expected. As soon as I know one variable, I begin substituting back into the other equations. Notice that I can find z by substituting $y = 1$ into equation 3. So, I get:

$$\begin{array}{r} x + y + z = 0 \\ y = 1 \\ 2 - z = 5 \end{array} \quad \longrightarrow \quad \begin{array}{r} x + 1 - 3 = 0 \\ y = 1 \\ z = -3 \end{array} \quad \longrightarrow \quad \begin{array}{r} x = 2 \\ y = 1 \\ z = -3 \end{array}$$

So, the solution to the system is the point (or vector in \mathbb{R}^3) $(2, 1, -3)$.

Now, let's talk about how I chose the steps that I took. It is important to notice that I made the choice from the start to eliminate all x 's below the first equation. So, I use the first equation to eliminate the others. I asked, "what can I multiply the first equation by so that when I add it to the second equation, the x will be eliminated?" The answer in this case was 2 and so I chose the rule $E_2 = -2e_1 + e_2$. I then ask a similar question for equation 3. The answer was 1 and so I chose the rule $E_3 = e_1 + e_3$. In the second step, I wanted to use the second equation to eliminate the y in the third equation (this was my plan). It's easier to do this if you have a coefficient of 1. I had a choice, I could have divided through by -3 , but that would have given me fractions. So, I used equation 3 to get my coefficient of 1. I noticed that $3 - 2 = 1$ so I chose the rule $E_2 = -e_2 + e_3$. It turned out that I was given a beautiful surprise because it eliminated z ! Notice that if I had a coefficient different than 1 for x in the first equation, I may have made different plans. Maybe I would rearrange my equations if there's another equation whose x coefficient is 1. Maybe I would choose to eliminate a different variable whose coefficient is 1 (or it's easier to get a coefficient of 1) in one of the equations. These are the things I look for.

Example 1.2. *Let's look at another system.*

$$\begin{array}{rcl}
 x + y + z = 2 & & x + y + z = 2 \\
 2x + 2y + 2z = 4 & \begin{array}{l} E_2 = -2e_1 + e_2 \\ \hline E_3 = e_1 + e_3 \end{array} & \begin{array}{l} 0 = 0 \\ 2y - z = 0 \end{array} \\
 -x + y - 2z = -2 & & \\
 & \begin{array}{l} E_1 = e_3 + e_1 \\ \hline E_2 = e_3, E_3 = e_2 \end{array} & \begin{array}{l} x + 3y = 2 \\ 2y - z = 0 \\ 0 = 0 \end{array}
 \end{array}$$

Notice that in the original system, the first two equations were the same plane. This became very apparent when I tried to eliminate x in the second equation. All variables eliminated. We know that at the least, two planes that intersect intersect at infinitely many points. This means that we will get either no solution or infinitely many solutions. We now back substitute to find all solutions: We let y be any real number and see that $z = 2y$ and $x = 2 - 3y$. So our solutions are points of the form $(2 - 3y, y, 2y)$, where y can be any real number. We say that y is a free variable in this case. Notice that we chose y to be the free variable. We could have chosen any of the variables in this problem to be free. Finally, since there are infinitely many solutions, we should write the solution set. The solution set is the set of all solutions to the system of equations. In this case the solution set is the set

$$\{(2 - 3y, y, 2y) \mid y \in \mathbb{R}\}.$$

In our first example, there is only one point in the solution set. So, we could (though we don't usually) write that the solution set for the first example is $\{(2, 1, -3)\}$.

We should also look at a system of equations that doesn't have a solution.

Example 1.3.

$$\begin{array}{rcl}
 x + y + z = 2 & & x + y + z = 2 \\
 2x + 2y + 2z = 1 & \begin{array}{l} E_2 = -2e_1 + e_2 \\ \hline E_3 = e_1 + e_3 \end{array} & \begin{array}{l} 0 = -3 \\ 2y - z = 0 \end{array} \\
 -x + y - 2z = -2 & &
 \end{array}$$

Notice that equation 2 is false. This means that there is not way for all three equations to be true at the same time. So, there is no solution. We would then write that the solutions set is \emptyset (the empty set).

1.3 Using Matrices to solve systems of equations

In this section, we introduce matrices to help with the bookkeeping. Later we will use matrices for other reasons and they may or may not be linked to systems of equations. What this means is that the use of matrices is more general (one might even say more powerful) than just solving systems of equations. With that said, let's talk about matrices. A matrix is a rectangular array of numbers. For example,

$$M = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 2 & -1 & 0.2 \end{pmatrix}$$

is a 2×4 matrix. (Note, some people use parenthesis and some people use square brackets around a matrix. Either is acceptable, and I may or may not go back and forth with these throughout the notes.) This means that M has two rows (horizontal) and four columns (vertical). Notice that a matrix has a nice structure where things are lined up, just like we lined things up when we wrote our systems of equations above. We use this structure to represent a system of equations as a matrix. In particular, we use what we call an augmented matrix. This only means that we put a vertical line in the matrix to keep track of where the equals sign is located.

For example, we represent the system of equations from the previous subsection

$$\begin{array}{rclcl} x & + & y & + & z & = & 0 \\ 2x & - & y & + & 3z & = & -6 \\ -x & + & y & - & 2z & = & 5 \end{array}$$

with the following coefficient matrix

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

Notice that it is called a coefficient matrix because the entries in the matrix are the coefficients from the system.

When solving systems of equations by hand, it can become very tedious writing the variables over and over. So, we use matrices to solve systems. In order to use a matrix, we need to note that when we eliminate, we are getting 0 entries. The process of elimination is now called matrix reduction. I use a similar notation as before, but now the equations are rows, so I use R and r instead of E and e to notate my steps. Let's solve the same system, but now using a matrix. Before we do, we rewrite the allowed operations in terms of rows:

Allowed operations when reducing a matrix.

- Multiply a row by a nonzero number.
- Add multiples of rows together to change only one of them.
- Change the order of rows.

Using these allowed operations, we will reduce the augmented matrix:

$$\begin{aligned} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 2 & -1 & 3 & -6 \\ -1 & 1 & -2 & 5 \end{array} \right) & \xrightarrow[\begin{array}{l} R_2 = -2r_1 + r_2 \\ R_3 = r_1 + r_3 \end{array}]{R_2 = -2r_1 + r_2} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & -3 & 1 & -6 \\ 0 & 2 & -1 & 5 \end{array} \right) \\ & \xrightarrow{R_2 = -r_2 - r_3} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 2 & -1 & 5 \end{array} \right). \end{aligned}$$

Here, we can write the system corresponding to this last matrix and then back substitute (which is just finishing up like we did in the last subsection) or we can continue to reduce. If we reduce so that the 2 in row three is a zero, we will have the matrix in what we call echelon form. If we continue to reduce so that down the main diagonal, there are 1's and above and below these 1's are 0's, we will have reduced to reduced echelon form. Let's take it all the way to reduced echelon form.

$$\begin{aligned} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 2 & -1 & 5 \end{array} \right) & \xrightarrow[\begin{array}{l} R_1 = -r_2 + r_1 \\ R_3 = -2r_2 + r_3 \end{array}]{R_1 = -r_2 + r_1} \left(\begin{array}{ccc|c} 1 & 0 & 1 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -1 & 3 \end{array} \right) \\ \xrightarrow[\begin{array}{l} R_3 = -r_3 \end{array}]{R_3 = -r_3} \left(\begin{array}{ccc|c} 1 & 0 & 1 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -3 \end{array} \right) & \xrightarrow{R_1 = -r_3 + r_1} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -3 \end{array} \right). \end{aligned}$$

Notice that the final matrix corresponds to the system $x = 2, y = 1, z = -3$. That is, the solution is $(2, 1, -3)$.

There's some terminology that we skimmed over above. Let's solidify the definitions.

Definition 1.1. Leading entries in a matrix are the first nonzero entries in a row, when reading from left to right.

Definition 1.2. Echelon form of a matrix is the obtained when the following three things are true:

- All leading entries are 1.
- All entries below each leading one is a zero.
- Any row of zeros is below a row that is not all zero.

Definition 1.3. Reduced echelon form of a matrix is obtained when the following three things are true:

- All leading entries are 1.
- All entries above and below each leading one is a zero.
- Any row of zeros is below a row that is not all zero.

Notice that the matrix on the left below is in echelon form while the matrix on the right is in reduced echelon form:

$$\begin{pmatrix} 1 & 3 & -2 & 5 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & -8 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Note also, that these matrices need not represent systems of equations.

Exercises

Don't forget to do the two exercises in the notes above.

1. Using both methods above, solve the system of equations. Verify your solution by plotting both equations.

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

2. Using both methods above, solve the system of equations. Verify your solution by plotting both equations.

$$\begin{aligned} 2x + 3y &= -5 \\ 2x - 2y &= 10 \end{aligned}.$$

3. Plot both equations in the system below. Solve the system of equations. What do you notice?

$$\begin{aligned} 2x + 3y &= -5 \\ 4x + 6y &= 10 \end{aligned}.$$

4. Plot both equations in the system below. Solve the system of equations. What do you notice?

$$\begin{aligned} x + 2y &= 5 \\ 2x + 4y &= 10 \end{aligned}.$$

5. Give an example of a system of linear equations with no solution. What must be true about the graphs of these equations?

6. Give an example of a system of linear equations with infinitely many solutions. What must be true about the graphs of these equations?

7. Solve the system of equations. Write the solution set.

$$\begin{aligned} x - y - z &= 4 \\ 2x - y + 3z &= 2 \\ -x + y - 2z &= -1 \end{aligned}.$$

8. Solve the system of equations. Write the solution set.

$$\begin{aligned} 2x - y - 3z &= 1 \\ 3x + y - 3z &= 4 \\ -2x + y + 2z &= -1 \end{aligned}.$$

9. Solve the system of equations. Write the solution set.

$$\begin{aligned}x - 2y - 3z &= 2 \\4x + y - 2z &= 8 . \\5x - y - 5z &= 10\end{aligned}$$

10. Solve the system of equations. Write the solution set.

$$\begin{aligned}x - 2y - 3z &= 2 \\4x + y - 2z &= 8 . \\5x - y - 5z &= 3\end{aligned}$$

11. A system of equations is called homogeneous if the right-hand-sides of each equation is 0. What is always a solution to a homogeneous system of equations?

12. We allowed three operations to be performed in finding the solution to a system of equations. For each allowed operation, state how you know that performing this operation does not change the solution to the original system.

13. Use a matrix to solve the system of equations by reducing the matrix to reduced echelon form, be sure to write the solution set.

$$\begin{aligned}x - y - z &= 4 \\2x - y + 3z &= 2 . \\-x + y - 2z &= -1\end{aligned}$$

14. Use a matrix to solve the system of equations by reducing the matrix to reduced echelon form, be sure to write the solution set.

$$\begin{aligned}2x - y - 3z &= 1 \\3x + y - 3z &= 4 . \\-2x + y + 2z &= -1\end{aligned}$$

15. Use a matrix to solve the system of equations by reducing the matrix to reduced echelon form, be sure to write the solution set.

$$\begin{aligned}x - 2y - 3z &= 2 \\4x + y - 2z &= 8 . \\5x - y - 5z &= 10\end{aligned}$$

16. Use a matrix to solve the system of equations by reducing the matrix to reduced echelon form, be sure to write the solution set.

$$\begin{aligned}x - 2y - 3z &= 2 \\4x + y - 2z &= 8 . \\5x - y - 5z &= 3\end{aligned}$$

2 Matrix operations and Matrix Equations

In case you have never used a matrix, seen a matrix, or played around with them at all, this section will give the rules and procedures for doing operations on and with matrices. If you are very familiar with these operations, this section will be mostly review. It happens that in the mathematical world, when we introduce a new entity, we tend to want to do things on or with it. In the case of matrices, we've seen that we can reduce them, but there is so much more.

To make the following easier to discuss, we introduce some standard notation. We say $M = (m_{i,j})$ and mean that M is a matrix made up of entries where the i, j th entry is $m_{i,j}$. That is, in the i th row and j th column is the number $m_{i,j}$. For example, let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 5 & 7 & 9 \end{pmatrix}.$$

Then, we can say $A = (a_{i,j})$ and this tells us that $a_{1,1} = 1$, $a_{1,2} = 2$, $a_{1,3} = 3$, $a_{2,1} = 5$, $a_{2,2} = 7$, and $a_{2,3} = 9$. Note that this notation is mostly a way of speaking about general results and definitions about matrices.

2.1 Basic Operations on Matrices

When working with matrices it becomes very important to know the dimensions of the matrix with which we are working. So, here we make this concept clearer. A matrix is an array of numbers arranged in a rectangular grid. We say that the dimensions of a matrix are m and n or that the matrix is an $m \times n$ matrix if the number of rows (horizontal) is m and the number of columns (vertical) is n . For example, the matrix

$$M_1 = \begin{pmatrix} 1 & -2 & 3 \\ 0 & 4 & 10 \\ 7 & 7 & 8 \\ -1 & 0 & 0 \end{pmatrix}$$

is 4×3 matrix, whereas

$$M_2 = \begin{pmatrix} 1 & 0 & 7 & -1 \\ -2 & 4 & 7 & 0 \\ 3 & 10 & 8 & 0 \end{pmatrix}$$

is a 3×4 matrix.

This example brings up an important operation on matrices.

Definition 2.1. *The transpose of a matrix M is to make the columns rows and the rows columns. If N is the transpose of M , we write $N = M^T$. (Note: The T is a superscript, not an exponent.)*

In the example above, if we take the transpose of M_1 , we get M_2 . That is, $M_2 = M_1^T$. But this means that $M_2^T = M_1$ also. So, the transpose can be undone, by applying the transpose twice.

2.1.1 Matrix Addition

Like many mathematical objects, we like to be able to combine them. The first way we ever learned to combine numbers was using addition. We will now define matrix addition.

Definition 2.2. Given two matrices $A = (a_{i,j})$ and $B = (b_{i,j})$. If they have the same dimension, then we define matrix addition by $A + B = (a_{i,j} + b_{i,j})$.

That is, we add the matrices component-wise as long as they have components that can be matched up. Notice that addition is not defined for two matrices that have different dimensions. For example, let

$$P = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}, \quad Q = \begin{pmatrix} 2 & 5 \\ 3 & -2 \\ 7 & -8 \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} 2 & 0 \\ 1 & -2 \end{pmatrix}.$$

Then

$$P + Q = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} + \begin{pmatrix} 2 & 5 \\ 3 & -2 \\ 7 & -8 \end{pmatrix} = \begin{pmatrix} 1+2 & 2+5 \\ 3+3 & 4-2 \\ 5+7 & 6-8 \end{pmatrix} = \begin{pmatrix} 3 & 7 \\ 6 & 2 \\ 12 & -2 \end{pmatrix}.$$

But neither $P + R$ nor $Q + R$ are defined.

2.1.2 Matrix Multiplication

Being able to add matrices encourages us to consider multiplication of matrices. There are several different ways one can multiply with a matrix. In this subsection, we will discuss a few of these.

Definition 2.3. Given a set of numbers, X , which we will call scalars, we define scalar multiplication in the following way. Let $\alpha \in X$ be a scalar and let $A = (a_{i,j})$ be a matrix. Then $\alpha A = (\alpha a_{i,j})$.

That is, we multiply each entry in P by the scalar α . For example, using P above, we get

$$2P = 2 \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 2 \cdot 1 & 2 \cdot 2 \\ 2 \cdot 3 & 2 \cdot 4 \\ 2 \cdot 5 & 2 \cdot 6 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 6 & 8 \\ 10 & 12 \end{pmatrix}.$$

Definition 2.4. Given two $m \times n$ matrices $A = (a_{i,j})$ and $B = (b_{i,j})$, we define the generalized dot product by $A \cdot B = \sum_{i=1}^m \sum_{j=1}^n a_{i,j} b_{i,j}$.

Note, this is the same rule as the dot product used when discussing vectors in a physics or vector Calculus class. For example, using P and Q above, we see that

$$P \cdot Q = 1 \cdot 2 + 2 \cdot 5 + 3 \cdot 3 + 4 \cdot (-2) + 5 \cdot 7 + 6 \cdot (-8) = 0.$$

Definition 2.5. Given an $n \times m$ matrix $A = (a_{i,j})$ and an $m \times \ell$ matrix $B = (b_{i,j})$, we define matrix multiply by $AB = (c_{i,j})$ where $c_{i,j}$ is obtained by finding the dot product between the i th row of A and the j th column of B .

Notice that there is no “.” between the matrices, rather they are written in juxtaposition to indicate a matrix product. Notice also that in order for a matrix product to be defined, the number of columns of the first matrix must match the number of rows of the second. For example, using P, Q , and R above, we see that PQ is not defined because both P and Q are 3×2 matrices. But, PR is defined. Let’s compute it.

$$PR = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 1 & -2 \end{pmatrix} = \begin{pmatrix} 1 \cdot 2 + 2 \cdot 1 & 1 \cdot 0 + 2 \cdot (-2) \\ 3 \cdot 2 + 4 \cdot 1 & 3 \cdot 0 + 4 \cdot (-2) \\ 5 \cdot 2 + 6 \cdot 1 & 5 \cdot 0 + 6 \cdot (-2) \end{pmatrix} = \begin{pmatrix} 4 & -4 \\ 10 & -8 \\ 16 & -12 \end{pmatrix}.$$

Let’s do another example. Let

$$P = \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 1 \end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix} 2 & 1 & -1 & 0 \\ 1 & -1 & 1 & 1 \\ 2 & 1 & 1 & 2 \end{pmatrix}.$$

Notice that P is 2×3 and Q is 3×4 . So, we can multiply them and the resulting matrix will be 2×4 . Indeed,

$$\begin{aligned} PQ &= \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & -1 & 0 \\ 1 & -1 & 1 & 1 \\ 2 & 1 & 1 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot 2 + 0 \cdot 1 + 2 \cdot 2 & 1 \cdot 1 + 0 \cdot (-1) + 2 \cdot 1 & 1 \cdot (-1) + 0 \cdot 1 + 2 \cdot 1 & 1 \cdot 0 + 0 \cdot 1 + 2 \cdot 2 \\ 2 \cdot 2 + (-1) \cdot 1 + 1 \cdot 2 & 2 \cdot 1 + (-1) \cdot (-1) + 1 \cdot 1 & 2 \cdot (-1) + (-1) \cdot 1 + 1 \cdot 1 & 2 \cdot 0 + (-1) \cdot 1 + 1 \cdot 2 \end{pmatrix} \\ &= \begin{pmatrix} 6 & 3 & 1 & 4 \\ 5 & 4 & -2 & 1 \end{pmatrix}. \end{aligned}$$

Notice that in neither case above, is the product QP defined. It turns out that matrix multiply is not commutative.

2.2 Matrix Equations and more

Now that we know about matrix multiply, we can consider a system of equations again using matrices. First, let us define what it means for two matrices to be equal.

Definition 2.6. *Given two matrices $A = (a_{i,j})$ and $B = (b_{i,j})$, we say that $A = B$ if and only if A and B have the same dimensions and $a_{i,j} = b_{i,j}$ for every pair i, j . That is, corresponding components of A and B are equal.*

Now, let us consider the system of linear equations from before:

$$\begin{aligned} x + y + z &= 0 \\ 2x - y + 3z &= -6 \\ -x + y - 2z &= 5 \end{aligned}$$

This system equations relates to the matrix equality given by

$$\begin{pmatrix} x + y + z \\ 2x - y + 3z \\ -x + y - 2z \end{pmatrix} = \begin{pmatrix} 0 \\ -6 \\ 5 \end{pmatrix}.$$

(Notice that each side of this equation is a 3×1 matrix.) Using what we know from matrix multiply, we can rewrite the left-hand-side as a matrix multiply. Doing this gives us

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & 3 \\ -1 & 1 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ -6 \\ 5 \end{pmatrix}.$$

If we let

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & 3 \\ -1 & 1 & -2 \end{pmatrix}, \quad X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \text{and} \quad b = \begin{pmatrix} 0 \\ -6 \\ 5 \end{pmatrix},$$

then this equation is just

$$AX = b.$$

This looks like a simple equation from an algebra class. From here we will discuss how we can solve this system of equations by “undoing the matrix multiply on the left side.” Note: we have not defined matrix division because it doesn’t exist. But, consider what division of real numbers really means. Let’s look at a simple example. Consider

$$2x = 4.$$

To solve this equation, we divide both sides by 2 and get $x = 2$. If we don’t allow division and only allow multiplication, we would just multiply both sides by $\frac{1}{2}$ instead and we get the same answer $x = 2$. So the availability of this multiplicative inverse allows us to remove the 2 from the left side of the equation. We want the same option with matrix equations. Notice that if we multiply a real number by it’s multiplicative inverse (which exists as long as the number is not 0), then we get the multiplicative identity, 1: $a \cdot \frac{1}{a} = \frac{1}{a} \cdot a = 1$. We want to be able to do the same with matrices when possible. First, we define the multiplicative identity matrix.

Definition 2.7. We define the $n \times n$ identity matrix, $I_{n \times n}$, as the square matrix with leading 1’s in every row that is in reduced echelon form.

Examples are for $n = 1$: (1), for $n = 2, 3$, and 4 are (respectively)

$$I_{2 \times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I_{3 \times 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{and} \quad I_{4 \times 4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

It is important to recognize why this is called the identity matrix. We know that 1 is the multiplicative identity for the real numbers because $a \cdot 1 = 1 \cdot a = a$ for any real number a . I claim that $I_{n \times n}$ is the multiplicative identity in the sense that $I_{n \times n}A = AI_{n \times n} = A$ for any $n \times n$ matrix A . Our goal is to find the multiplicative inverse (or simply called, the inverse) of a matrix when it exists. Let us now define what we want.

Definition 2.8. Let A be an $n \times n$ matrix, we say B is the inverse of A if and only if B is an $n \times n$ matrix so that

$$AB = BA = I_{n \times n}.$$

If B satisfies this definition, then we say that A is invertible and we write

$$B = A^{-1}.$$

(Note that the -1 is a superscript not an exponent.)

Let's now spend some time exploring how we compute (or find) the inverse. Suppose we have a matrix

$$B = \left(\begin{array}{c|c|c|c} | & | & & | \\ b_1 & b_2 & \dots & b_n \\ | & | & & | \end{array} \right)$$

with columns b_1, b_2, \dots, b_n . Suppose also that B is the inverse of A . Then since $AB = I_{n \times n}$, we know that $Ab_1 = e_1$, where e_1 is the first column of $I_{n \times n}$, $Ab_2 = e_2$, where e_2 is the second column of $I_{n \times n}$, and so on. So all we need to do is solve the n matrix equations:

$$\begin{aligned} Ab_1 &= e_1 \\ Ab_2 &= e_2 \\ &\vdots \\ Ab_n &= e_n. \end{aligned}$$

Let's look at an example to see how this can be done. We'll start with a simple example. Let

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}.$$

We want to find A^{-1} . That is, we want to find the matrix

$$B = \left(\begin{array}{c|c} | & | \\ b_1 & b_2 \\ | & | \end{array} \right)$$

so that $AB = I_{2 \times 2}$. We find it one column at a time by solving the equations

$$\begin{aligned} \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} b_{1,1} \\ b_{2,1} \end{pmatrix} &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} b_{1,2} \\ b_{2,2} \end{pmatrix} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned}$$

Notice that the first equation is the same as solving the system of equations

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

And the second equation is the same as solving the system of equations

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

We can solve these by using matrix reduction on the two augmented matrices

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

and

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

Following the same procedure, we augment A with $I_{3 \times 4}$ and reduce.

$$\begin{array}{l}
 \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 2 & -1 & 3 & 0 & 1 & 0 \\ -1 & 1 & -2 & 0 & 0 & 1 \end{array} \right) \xrightarrow{\substack{R_2 = -2r_1 + r_2 \\ R_3 = r_1 + r_3}} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & -3 & 1 & -2 & 1 & 0 \\ 0 & 2 & -1 & 1 & 0 & 1 \end{array} \right) \\
 \xrightarrow{R_2 = -r_3 - r_2} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 & -1 \\ 0 & 2 & -1 & 1 & 0 & 1 \end{array} \right) \xrightarrow{\substack{R_1 = -r_2 + r_1 \\ R_3 = -2r_2 + r_3}} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & -1 & -1 \\ 0 & 0 & -1 & -1 & 2 & 3 \end{array} \right) \\
 \xrightarrow{\substack{R_1 = r_3 + r_1 \\ R_3 = -r_3}} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 3 & 4 \\ 0 & 1 & 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 1 & -2 & -3 \end{array} \right).
 \end{array}$$

Thus,

$$A^{-1} = \begin{pmatrix} -1 & 3 & 4 \\ 1 & -1 & -1 \\ 1 & -2 & -3 \end{pmatrix}.$$

We then use this to solve $AX = b$ by multiplying both sides of this equation on the left by A^{-1} . In doing this we get the following string of equations:

$$AX = b \Rightarrow A^{-1}AX = A^{-1}b \Rightarrow IX = A^{-1}b \Rightarrow X = A^{-1}b.$$

(Here for simplicity, I used I for $I_{3 \times 3}$.) So, we can find our solution by multiplying our vector on the right-hand-side by A^{-1} . That is,

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

This is the solution we found the last two times we solved this same system, so that feels pretty good. In the exercises, you will explore this further.

2.3 Determinants

In this section, we discuss how to calculate a determinant of a matrix and discuss some information that we can obtain by finding the determinant. First, note that this section is not a comprehensive discussion of the determinant of a matrix. There are geometric interpretations that we will not discuss here.

This section will present the computation and then help you put together why we would ever consider finding the determinant. At first, this computation seems lengthy and maybe even more work than its worth. But, if you hold on until the end, you will see that the determinant is a very useful tool. As a hint, we will use this tool to help us decide the outcome when solving systems of equations or when solving matrix equations. We know the solutions to easy matrix equations such as

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

or

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

But there are so many scenarios in between for which we would like to know a hint about the solution before we begin the journey through matrix reduction (especially if we have to do these by hand and if they are big). With that, we begin by discussing the properties that we want the determinant of a matrix to have. These properties are all related to matrix reduction steps.

Let α be a scalar. We want the determinant of an $n \times n$ matrix, A , to satisfy the following properties

- $\det(\alpha A) = \alpha^n \det(A)$.
- $\det(A^T) = \det(A)$.
- If B is obtained by performing the row operation, $R_k = \alpha r_j + r_k$ on A , then $\det(B) = \det(A)$.
- If B is obtained by performing the row operation, $R_k = \alpha r_k$ on A , then $\det(B) = \alpha \det(A)$.
- If B is obtained by performing the row operation, $R_k = r_j$ and $R_j = r_k$ on A , then $\det(B) = -1 \cdot \det(A)$.
- If A is in echelon form, then $\det(A)$ is the product of the diagonal elements.

We can use these properties to find the determinant of a matrix by keeping track of the determinant as we perform row operations on the matrix. Let us try an example. We will find the determinant of

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & 3 \\ -1 & 1 & -2 \end{pmatrix}.$$

Our goal is to reduce A to echelon form all the while keeping track of how the determinant changes.

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & 3 \\ -1 & 1 & -2 \end{pmatrix} \xrightarrow[\begin{matrix} R_2 = -2r_1 + r_2 \\ R_3 = r_1 + r_3 \end{matrix}]{\det(A)} \begin{pmatrix} 1 & 1 & 1 \\ 0 & -3 & 1 \\ 0 & 2 & -1 \end{pmatrix} \xrightarrow[\det(A)]{R_2 = r_2 + r_3} \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 2 & -1 \end{pmatrix} \det(A) \\ \xrightarrow[\begin{matrix} R_2 = -r_2 \\ R_3 = 2r_2 + r_3 \end{matrix}]{- \det(A)} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \xrightarrow[\det(A)]{R_3 = -r_3} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \det(A).$$

Now to use the last property above, we see that

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

Thus, $\det(A) = 1$.

We will try one more example before giving another method for finding the determinant. We will find the determinant of

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

Again, we will reduce A to echelon form all the while keeping track of how the determinant changes.

$$\begin{array}{ccc}
 \begin{pmatrix} 2 & 2 & 2 \\ 1 & 0 & 1 \\ -2 & 2 & -4 \end{pmatrix} & \xrightarrow{R_1 = \frac{1}{2}r_1} & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ -2 & 2 & -4 \end{pmatrix} & \xrightarrow{\substack{R_2 = -r_1 + r_2 \\ R_3 = 2r_1 + r_3}} & \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 4 & -2 \end{pmatrix} \\
 \det(A) & & \frac{1}{2} \det(A) & & \frac{1}{2} \det(A) \\
 \\
 \begin{array}{ccc}
 \xrightarrow{\substack{R_2 = -r_2 \\ R_3 = 4r_2 + r_3}} & \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} & \xrightarrow{R_3 = -\frac{1}{2}r_3} & \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
 & -\frac{1}{2} \det(A) & & \frac{1}{4} \det A
 \end{array}
 \end{array}$$

Now to use the last property above, we see that

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

Thus, $\frac{1}{4} \det(A) = 1$ and thus, $\det(A) = 4$.

Clearly, there has to be another method because, well, I said that we would want to know the determinant before going through all of those steps. Another method for finding the determinant of a matrix is the method called cofactor expansion. First, if the matrix M is 2×2 , it is much easier to compute the determinant. Here's the formula: Let

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then the determinant is $\det M = ad - bc$. (This can be shown using the above steps, see exercise below.) Note: $\det M$ is also written as $|M|$ and from here on, we will use this notation because it is simpler.

If M is a bigger matrix, then there's more to do here. Here, we write out the formula given by this method.

Let $A = (a_{i,j})$ be an $n \times n$ matrix and choose any j so that $1 \leq j \leq n$, then

$$|A| = \sum_{i=1}^n (-1)^{i+j} a_{i,j} |M_{i,j}|,$$

where $M_{i,j}$ is the sub-matrix of A where the i th row and j th column has been removed. We can also expand about a column so that we choose an i so that $1 \leq i \leq n$, then

$$|A| = \sum_{i=j}^n (-1)^{i+j} a_{i,j} |M_{i,j}|,$$

Notice that if n is large, this process is iterative until the sub-matrices are 2×2 . Here are couple examples showing what this formula looks like.

Example 2.1. First, we consider an example when $n = 3$. Let $A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}$,

then

$$|A| = a_{1,1} \begin{vmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{vmatrix} - a_{1,2} \begin{vmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{vmatrix} + a_{1,3} \begin{vmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{vmatrix}.$$

From there, you just use the formula for the determinant of a 2×2 to finish this.

Example 2.2. In the previous example, we expanded about the first row. Here, we consider another example when $n = 3$, but now expand about the second column. Let $A =$

$$\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}, \text{ then}$$

$$|A| = -a_{1,2} \begin{vmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{vmatrix} + a_{2,2} \begin{vmatrix} a_{1,1} & a_{1,3} \\ a_{3,1} & a_{3,3} \end{vmatrix} - a_{3,2} \begin{vmatrix} a_{1,1} & a_{1,3} \\ a_{2,1} & a_{2,3} \end{vmatrix}.$$

Again, you finish this by employing the formula for the determinant of 2×2 matrices.

Example 2.3. Let $A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix}$, then

$$|A| = a_{1,1} \begin{vmatrix} a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,2} & a_{4,3} & a_{4,4} \end{vmatrix} - a_{1,2} \begin{vmatrix} a_{2,1} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,3} & a_{4,4} \end{vmatrix} \\ + a_{1,3} \begin{vmatrix} a_{2,1} & a_{2,2} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,4} \end{vmatrix} - a_{1,4} \begin{vmatrix} a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \\ a_{4,1} & a_{4,2} & a_{4,3} \end{vmatrix}.$$

From here, you need to employ a formula for the determinant of 3×3 matrices (as in Example 1 or 2 above).

Now, we shall apply this technique to the matrix:

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

We will expand about the second column (because there is a 0, it'll be less to keep track of).

$$|A| = -2 \begin{vmatrix} 1 & 1 \\ -2 & -4 \end{vmatrix} + 0 \begin{vmatrix} 2 & 2 \\ -2 & -4 \end{vmatrix} - 2 \begin{vmatrix} 2 & 2 \\ 1 & 1 \end{vmatrix} = -2(-4 - (-2)) + 0 - 2(2 - 2) = 4.$$

Exercises

1. For each of the exercises below, use the matrices

$$P = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 2 & 1 & 3 & -1 \\ 2 & 2 & 0 & 0 \end{pmatrix}, \quad Q = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \\ 1 & 2 & 0 \\ 3 & 0 & 0 \\ 2 & 2 & 2 \\ 1 & 0 & -1 \end{pmatrix}, \quad R = \begin{pmatrix} 0 & 2 & 3 & 4 \\ 1 & -1 & -3 & -1 \\ -1 & -2 & 0 & 1 \end{pmatrix},$$

$$\text{and } S = \begin{pmatrix} 0 & 0 & -1 & 0 & 2 & 1 \\ -3 & -1 & -2 & 1 & 2 & 0 \\ 2 & -1 & 0 & 1 & 2 & -1 \end{pmatrix}$$

- (a) Compute $P + Q$, $R + P$, $P + R$, and $Q + S$. If the addition is not defined, state how you know.
- (b) Compute $P \cdot Q$, $Q \cdot R$, and $Q \cdot S$. If the generalized dot product is not defined, state how you know.
- (c) Compute $3P$, $2Q$, and $3S$. If the scalar product is not defined, state how you know.
- (d) Compute PQ , QP , QS , QR , RS and PR . If the matrix product is not defined, state how you know.
- (e) Compute P^T , Q^T , and RP^T .
2. Multiply $I_{2 \times 2}$ by a general 2×2 matrix, $A = (a_{i,j})$ to verify $I_{2 \times 2}$ is truly a multiplicative identity.
3. Multiply $I_{3 \times 3}$ by a general 3×3 matrix, $A = (a_{i,j})$ to verify $I_{3 \times 3}$ is truly a multiplicative identity.
4. Use the method of this section to find the inverse of

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

5. Based on what you notice happens when multiplying by $I_{n \times n}$ (from exercises 1 and 2), give a word justification that $I_{n \times n}$ is indeed a multiplicative identity.
6. Rewrite the system of equations as a matrix equation. Find the inverse of the 3×3 matrix on the left side (if possible). If the inverse exists, use the inverse to find the solution to the system. If not, state how you know. How many solutions does this system have? (This system is the same as was given in a previous exercise set.)

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

7. Rewrite the system of equations as a matrix equation. Find the inverse of the 3×3 matrix on the left side (if possible). If the inverse exists, use the inverse to find the solution to the system. If not, state how you know. How many solutions does this system have? (This system is the same as was given in a previous exercise set.)

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

8. Rewrite the system of equations as a matrix equation. Find the inverse of the 3×3 matrix on the left side (if possible). If the inverse exists, use the inverse to find the solution to the system. If not, state how you know. How many solutions does this system have? (This system is the same as was given in a previous exercise set.)

$$\begin{aligned} x - 2y - 3z &= 2 \\ 4x + y - 2z &= 8 \\ 5x - y - 5z &= 10 \end{aligned}$$

9. Rewrite the system of equations as a matrix equation. Find the inverse of the 3×3 matrix on the left side (if possible). If the inverse exists, use the inverse to find the solution to the system. If not, state how you know. How many solutions does this system have? (This system is the same as was given in a previous exercise set.)

$$\begin{array}{rcl} x & - & 2y & - & 3z & = & 2 \\ 4x & + & y & - & 2z & = & 8 \\ 5x & - & y & - & 5z & = & 3 \end{array}$$

10. Recall that a system of equations is called homogeneous if the right-hand-sides of each equation is 0. How many solutions does a homogeneous system have if the coefficient matrix is not invertible?
11. What can you say about the number of solutions to a system of equations if the inverse of the coefficient matrix exists? What if it does not exist?
12. Use the properties of determinant to verify the formula for the determinant of a 2×2 matrix. That is, reduce

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

keeping track of what happens to the determinant.

For the next six exercises, use any of the above methods (it's good to try multiple methods to make an informed decision here) to find the following determinants.

13. $\begin{vmatrix} 1 & 3 \\ 0 & 1 \end{vmatrix}$

14. $\begin{vmatrix} 2 & -3 \\ 1 & 1 \end{vmatrix}$

15. The determinant of the corresponding coefficient matrices from exercise 6 above.
16. The determinant of the corresponding coefficient matrices from exercise 7 above.
17. The determinant of the corresponding coefficient matrices from exercise 8 above.
18. The determinant of the corresponding coefficient matrices from exercise 9 above.
19. Using the previous 4 exercises, state what you notice about the determinant compared to whether or not the matrix is invertible.
20. The invertible matrix theorem is an important theorem. Fill in the blanks or circle the correct answer below to complete the statement of the theorem.
- (a) $AX = b$ has a unique solution if _____
(Choose one: A is invertible or A is not invertible).
- (b) A is invertible if and only if $\det A$ _____.
- (c) $AX = b$ has a unique solution if $\det A$ _____.

3 Radiography and Tomography in Linear Algebra Lab #1

This lab will be discussed in class. We will introduce the application with a visual describing our overall goal for this application. For this lab, you will need a computer with Matlab or Octave on it. An alternative option for this lab, but not all labs in this class would be to use Octave-online.

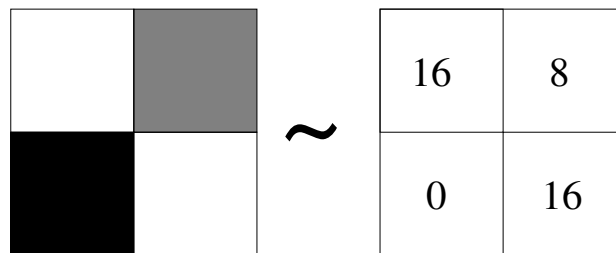
Instructions for using octave-online.net

1. Open octave-online.net in a web browser. Close (or read and close) the introduction splash window.
2. Optional: Click the three bars in the upper right of the screen and sign in. This will be necessary if you complete the next Lab using this service.
3. The rightmost window is the command prompt terminal. You will type your commands here. You also have the option of creating your own scripts and functions to save and run.
4. Now you are ready to complete the assignment!

3.1 Grayscale Images

A grayscale image can be associated with a set of numbers representing the brightness of each pixel.

For example, the 2×2 pixel image below can be associated with the corresponding array of numbers, where the black squares represent pixels with intensity 0 and the white squares represent pixels with intensity 16:



Note that this interpretation gives us a way to think about multiplying an image by a constant and adding two images. For example, suppose we start with the three images (A, B, and C) below.

Then multiplying Image A by 0.5 results in Image 1 below. Note that the maximum intensity is now half what it previously was, so the all pixels have become darker gray (representing their lower intensity). Adding Image 1 to Image C results in Image 2 below; so Image 2 is created by doing arithmetic on Images A and C.

Exercise 1: Write Image 3 and Image 4 using arithmetic operations of Images A, B, and C.

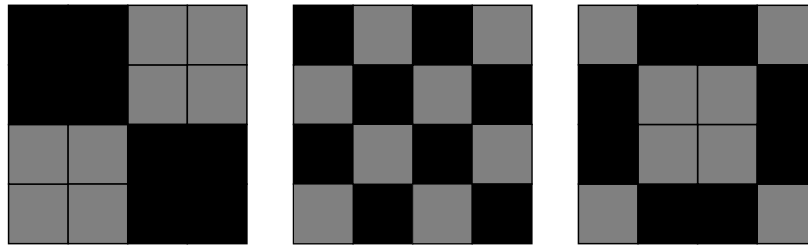


Image A

Image B

Image C

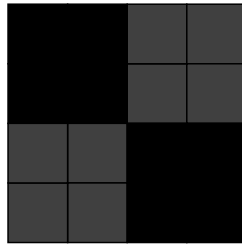


Image 1

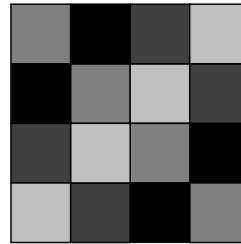


Image 2

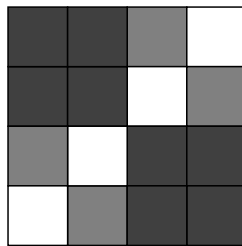


Image 3

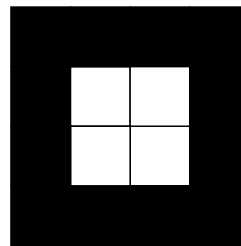


Image 4

Images in Matlab

We'll now see how to input matrices, convert them into grayscale images, and display them in Matlab (or Octave). If you do not have access to Matlab or Octave through academic or personal computing resources then you may complete this Lab using the online service octave-online.net (instructions at the end of this writeup).

Exercise 2 Input these lines into the command window of Matlab. Note that ending a line with a semicolon suppresses the output. If you want to show the result of a computation, delete the semicolon at the end of its line. Briefly describe what the output of each of these lines of code gives you. Note: In order to display an image, we have to know the minimum and maximum pixel intensities. The `imshow` function allows the user to specify these values.

```
M_A = [0 0 8 8; 0 0 8 8; 8 8 0 0; 8 8 0 0];
M_B = [0 8 0 8; 8 0 8 0; 0 8 0 8; 8 0 8 0];
M_C = [8 0 0 8; 0 8 8 0; 0 8 8 0; 8 0 0 8];
figure;
subplot(1,3,1), imshow(M_A, [0, 16]),title('Image A');
subplot(1,3,2), imshow(M_B, [0, 16]),title('Image B');
subplot(1,3,3), imshow(M_C, [0, 16]),title('Image C');
```

Exercise 3: Enter the following lines of code one at a time and state what each one does.

```
M_A
M_1 = .5*M_A
M_2 = M_1 + M_C
figure;
subplot(1,2,1), imshow(M_1, [0, 16]), title('Image 1');
subplot(1,2,2), imshow(M_2, [0, 16]), title('Image 2');
```

Exercise 4 Now write your own lines of code to check your conjectures to get Images 3 and 4. How close are these to Images 3 and 4? Be sure to print out your code and submit it with this lab.

Food For Thought Be prepared to discuss these in class:

1. What happens when pixel intensities in an image exceed the display range as input into the `imshow` function?
2. How should we interpret pixel intensities that lie outside our specified range?
3. What algebraic properties does the set of all images have in common with the set of real numbers?

4 Vector Spaces

In Lab 1, we noticed that if you add two images, you get a new image. We also saw that if you multiply an image by a scalar, you get a new image. Let's recall the discussion in class. We saw in Lab 1 that (rectangular pixelated) images can be represented as a rectangular array of values or equivalently as a rectangular array of grayscale patches. This is a very natural idea especially since the advent of digital photography. It is tempting to consider an image (or image data) as a matrix – after all, it certainly looks like one. In our discussion, we defined an image in the following way.

Definition 4.1. An image is a finite ordered list of real values with an associated geometric array description.

Three examples of arrays along with an index system specifying the order of patches can be seen in Figure 1. Each patch would also have a numerical value indicating the brightness of the patch (not shown). The first is a regular pixel array commonly used for digital photography. The second is a hexagon pattern which also nicely tiles a plane. The third is a square pixel set with enhanced resolution toward the center of the field of interest. The key point here is that only the first example can be written as a matrix, but all satisfy the definition of image.

In recognizing the definition of an image, we were able to define the operations of scalar multiplication and image addition in a way that was similar to the same operations on matrices.

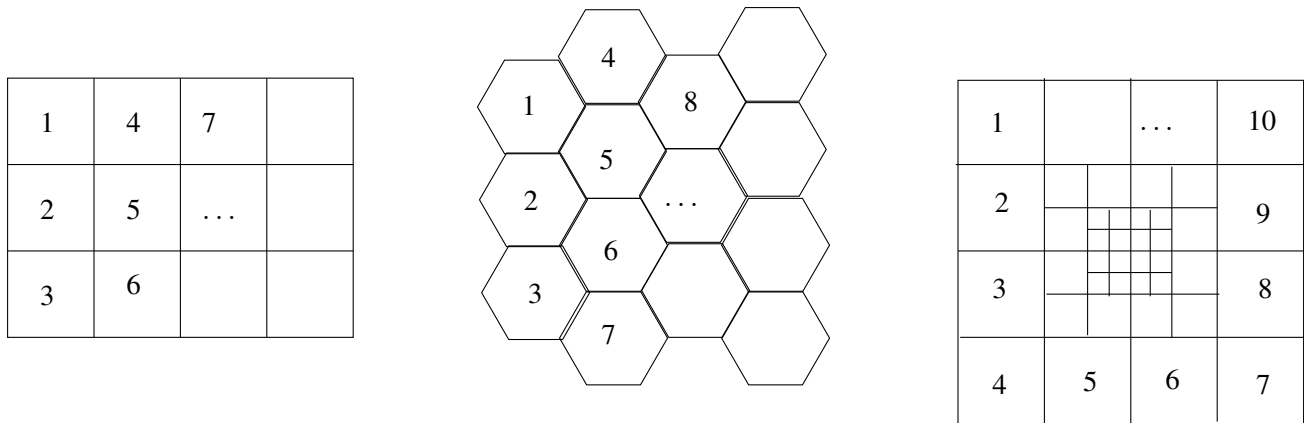


Figure 1: Examples of image arrays. Numbers indicate example pixel ordering.

Definition 4.2. Given two images x and y with (ordered) intensity values (x_1, x_2, \dots, x_n) and (y_1, y_2, \dots, y_n) , respectively, the image sum, written $z = x + y$ is the image with intensity values $z_i = x_i + y_i$ for all $i \in \{1, 2, \dots, n\}$. Another way to say this is that the sum of two images is the image that results by pixel-wise addition, that is, the sum of two images is the image that results by adding corresponding values of the ordered list.

Definition 4.3. Given scalar a and image x with (ordered) intensity values (x_1, x_2, \dots, x_n) , the scalar product of an image, written $z = ax$ is the image with intensity values $z_i = ax_i$ for all $i \in \{1, 2, \dots, n\}$. Another way to say this is that a scalar times an image is the image that results by pixel-wise scalar multiplication. That is, a scalar times an image is the image which results from multiplication of each of the intensity values (or values in the ordered list) by that scalar.

We found that these operations had a place in a real-life example and when applying these operations it is very important that our result is still an image in the same configuration. It turns out that these operations are important enough that we give a name to sets that have these operations with some pretty standard properties.

4.1 Vectors and Vector Spaces

The word vector may be familiar for many students who have taken Vector Calculus and/or Physics. In these courses, there is a very specific type of vector used, vectors in \mathbb{R}^m . That is, the word vector may bring to mind something that looks like $\langle a, b \rangle$, $\langle a, b, c \rangle$, or $\langle a_1, a_2, \dots, a_n \rangle$. Maybe you've even seen things like any of the following

$$(a, b), \quad (a, b, c), \quad (a_1, a_2, \dots, a_n), \quad \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \quad \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \quad \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

called vectors. In this subsection, we define the type of set we want to consider when talking about linear algebra concepts. Then, we discuss the elements, called vectors, of these sets.

Definition 4.4. A set V with a set of scalars and operations vector addition $+$ and scalar multiplication \cdot is called a vector space if the following ten properties hold. Elements of V are called vectors. Let $u, v, w \in V$ be vectors and α, β be scalars.

1. V is closed under addition $+$: $u + v \in V$. (Adding two vectors gives a vector.)
2. V is closed under scalar multiplication \cdot : $\alpha \cdot v \in V$. (Multiplying a vector by a scalar gives a vector.)
3. Addition is commutative: $u + v = v + u$.
4. Addition is associative: $(u + v) + w = u + (v + w)$.
5. Scalar multiplication is associative: $\alpha \cdot (\beta \cdot v) = (\alpha\beta) \cdot v$.
6. Scalar multiplication distributes over vector addition: $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$.
7. Scalar multiplication distributes over scalar addition: $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$.
8. V contains the additive identity or the 0 vector, where $0 + v = v + 0 = v$.
9. V has additive inverses $-v$: $v + (-v) = 0$.
10. The scalar set has an identity element 1 for scalar multiplication: $1 \cdot v = v$ for all $v \in V$.

It is important to note that the identity element for scalar multiplication need not be the number 1 and the zero vector need not be (and in general is not) the number 0.

Notice also that elements of a vector space are called vectors. These need not look like the vectors presented above.

We now present some examples of vector spaces and their corresponding vectors.

Example 4.1. \mathbb{R} , the set of real numbers is a vector space with scalars taken from the set of real numbers. By definition of addition $+$ between real numbers and multiplication \cdot between real numbers, they have all the properties above. So, there is nothing to show.

Here the vectors in this space are real numbers such as $0, 1, \sqrt{2}, \pi, \frac{10}{17}, 100$, and many more.

Example 4.2. \mathbb{R}^n , the set of points or vectors in n -dimensional space, is a vector space with scalars taken from the set of real numbers. Here, we recognize that addition of two vectors in \mathbb{R}^n is computed component-wise. We will show that \mathbb{R}^2 is a vector space and recognize

how the same proofs generalize. Let $\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \in \mathbb{R}^2$ be vectors and α, β be scalars.

- V is closed under addition $+$:

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \end{pmatrix} \in \mathbb{R}^2.$$

- V is closed under scalar multiplication \cdot :

$$\alpha \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \end{pmatrix} \in \mathbb{R}^2.$$

- *Addition is commutative:*

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \end{pmatrix} = \begin{pmatrix} v_1 + u_1 \\ v_2 + u_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

- *Addition is associative:*

$$\begin{aligned} \left(\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right) + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} &= \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \\ &= \begin{pmatrix} (u_1 + v_1) + w_1 \\ (u_2 + v_2) + w_2 \end{pmatrix} \\ &= \begin{pmatrix} u_1 + (v_1 + w_1) \\ u_2 + (v_2 + w_2) \end{pmatrix} \\ &= \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \end{pmatrix} \\ &= \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \left(\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \right). \end{aligned}$$

- *Scalar multiplication is associative:*

$$\alpha \cdot \left(\beta \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right) = \alpha \cdot \begin{pmatrix} \beta v_1 \\ \beta v_2 \end{pmatrix} = \begin{pmatrix} \alpha(\beta v_1) \\ \alpha(\beta v_2) \end{pmatrix} = \begin{pmatrix} (\alpha\beta)v_1 \\ (\alpha\beta)v_2 \end{pmatrix} = (\alpha\beta) \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

- *Scalar multiplication distributes over vector addition:*

$$\begin{aligned} \alpha \cdot \left(\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right) &= \alpha \cdot \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \end{pmatrix} \\ &= \begin{pmatrix} \alpha(u_1 + v_1) \\ \alpha(u_2 + v_2) \end{pmatrix} \\ &= \begin{pmatrix} \alpha u_1 + \alpha v_1 \\ \alpha u_2 + \alpha v_2 \end{pmatrix} \\ &= \begin{pmatrix} \alpha u_1 \\ \alpha u_2 \end{pmatrix} + \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \end{pmatrix} \\ &= \alpha \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \alpha \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}. \end{aligned}$$

- *Scalar multiplication distributes over scalar addition:*

$$\begin{aligned} (\alpha + \beta) \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} &= \begin{pmatrix} (\alpha + \beta)v_1 \\ (\alpha + \beta)v_2 \end{pmatrix} \\ &= \begin{pmatrix} \alpha v_1 + \beta v_1 \\ \alpha v_2 + \beta v_2 \end{pmatrix} \\ &= \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \end{pmatrix} + \begin{pmatrix} \beta v_1 \\ \beta v_2 \end{pmatrix} \\ &= \alpha \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \beta \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}. \end{aligned}$$

- V contains the 0 vector, where

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 + v_1 \\ 0 + v_2 \end{pmatrix} = \begin{pmatrix} v_1 + 0 \\ v_2 + 0 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

- V has additive inverses $-v$:

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} -v_1 \\ -v_2 \end{pmatrix} = \begin{pmatrix} v_1 + (-v_1) \\ v_2 + (-v_2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

- The scalar set has an identity element 1 for scalar multiplication:

$$1 \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1(v_1) \\ 1(v_2) \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

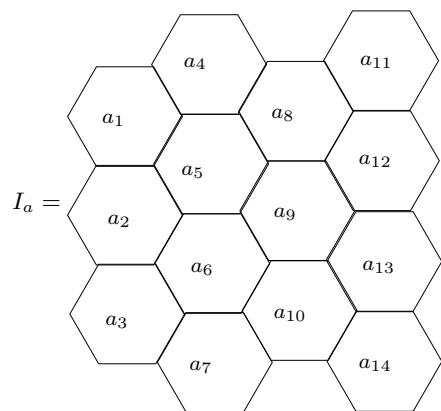
Here the vectors are just like the vectors we knew before Linear Algebra. Examples are

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} \pi \\ \sqrt{2} \end{pmatrix},$$

and many more.

Example 4.3. Notice that the set of images with a specified geometric arrangement is a vector space with scalars taken from \mathbb{R} . That is, if we consider the set

$$V = \{I_a \mid I \text{ is of the form below and } a_1, a_2, \dots, a_{14} \in \mathbb{R}\}.$$



We know this is a vector space since, by definition of addition and scalar multiplication on images, we see that both closure properties hold. Notice that there is a corresponding real number to each pixel (or voxel). Because addition and scalar multiplication are taken pixel-wise (or voxel-wise), we can see that these 10 properties hold within each pixel (or voxel). So, we know all 10 properties hold (just like they did in the last example).

It turns out that given any geometric configuration our definitions of image and operations on images guarantee that the space of images with the chosen configuration is a vector space. The vectors are then images.

Example 4.4. We can consider now the set of solutions to a system of equations. That is, consider

$$V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \mid x + y + z = 0, 2x - y + 3z = -6, \text{ and } -x + y - 2z = 5 \right\}.$$

We solved the system of equations

$$\begin{aligned} x + y + z &= 0 \\ 2x - y + 3z &= -6 \\ -x + y - 2z &= 5 \end{aligned}$$

several times above. We know the solution is $(2, 1, -3)$. Thus, the set $V = \left\{ \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} \right\}$.

Notice that V is not a vector space because it is not closed under scalar multiplication (nor is it closed under addition). Indeed, $\begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}$ is the only element of V and so no scalar multiple of this vector is in V .

What is important about the last example is that all we need to show, in order to show a set is not a vector space, is that one of the 10 properties fails.

Example 4.5. Let $\mathcal{F} = \{f : \mathbb{R} \rightarrow \mathbb{R}\}$, the set of all functions whose domain is \mathbb{R} and whose range is a subset of \mathbb{R} . \mathcal{F} is a vector space with scalars taken from \mathbb{R} . We can define addition and scalar multiplication in the standard way. That is, $f + g$ is the function so that $(f+g)(x) = f(x)+g(x)$ (adding point-wise) and αf is the function so that $(\alpha f)(x) = \alpha \cdot (f(x))$ (multiplying point-wise). Let $f, g, h \in \mathcal{F}$ and $\alpha, \beta \in \mathbb{R}$. Then:

- $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$. Based on the definition of addition, $f + g : \mathbb{R} \rightarrow \mathbb{R}$. So \mathcal{F} is closed over addition.
- Similarly, \mathcal{F} is closed under scalar multiplication.
- Addition is commutative:

$$(f + g)(x) = f(x) + g(x) = g(x) + f(x) = (g + f)(x).$$

So, $f + g = g + f$.

- Addition is associative:

$$\begin{aligned} ((f + g) + h)(x) &= (f + g)(x) + h(x) = (f(x) + g(x)) + h(x) = f(x) + (g(x) + h(x)) \\ &= f(x) + (g + h)(x) = (f + (g + h))(x). \end{aligned}$$

So $(f + g) + h = f + (g + h)$.

- Scalar multiplication is associative:

$$(\alpha \cdot (\beta \cdot f))(x) = (\alpha \cdot (\beta f(x))) = (\alpha\beta)f(x) = ((\alpha\beta) \cdot f)(x).$$

So $\alpha \cdot (\beta \cdot f) = (\alpha\beta) \cdot f$.

- *Scalar multiplication distributes over vector addition:*

$$\begin{aligned}(\alpha \cdot (f + g))(x) &= \alpha \cdot (f + g)(x) = \alpha \cdot (f(x) + g(x)) \\ &= \alpha \cdot f(x) + \alpha \cdot g(x) = (\alpha \cdot f + \alpha \cdot g)(x).\end{aligned}$$

So $\alpha(f + g) = \alpha f + \alpha g$.

- *Scalar multiplication distributes over scalar addition:*

$$\begin{aligned}((\alpha + \beta) \cdot f)(x) &= (\alpha + \beta) \cdot f(x) = \alpha \cdot f(x) + \beta \cdot f(x) \\ &= (\alpha \cdot f + \beta \cdot f)(x).\end{aligned}$$

So, $(\alpha + \beta) \cdot f = \alpha \cdot f + \beta \cdot f$.

- \mathcal{F} contains the constant function defined by $z(x) = 0$ for every $x \in \mathbb{R}$. And,

$$(z + f)(x) = z(x) + f(x) = 0 + f(x) = f(x) = f(x) + 0 = f(x) + z(x) = (f + z)(x).$$

That is, $z + f = f + z = f$. So, the 0 vector is in \mathcal{F} .

- \mathcal{F} has additive inverses $-f$ defined to as $(-f)(x) = -f(x)$ and $(f + (-f))(x) = f(x) + (-f(x)) = 0 = z(x)$, where z is defined in part 4.5. So, $f + (-f) = z$.
- The real number 1 satisfies: $(1 \cdot f)(x) = 1 \cdot f(x) = f(x)$. So, $1 \cdot f = f$.

Here, the vectors are functions.

4.2 Subspaces

Many times, we work in subsets of vector spaces. Because many of the ten properties don't go away when we look inside a set contained in a larger set, we need not reprove these properties. In this section we discuss subspaces.

Definition 4.5. Let V be a vector space. If $W \subseteq V$ (a subset of V), then we say that W is a subspace of V if it satisfies the following two properties.

1. $0 \in W$ (We don't allow W to be empty and so we require that it contains the 0 vector.)
2. Let $u, v \in W$ and α, β be scalars. Then $\alpha u + \beta v \in W$.

The last property is called the closure property under linear combinations. It makes sense to define linear combinations now.

Definition 4.6. A linear combination of vectors v_1, v_2, \dots, v_n is

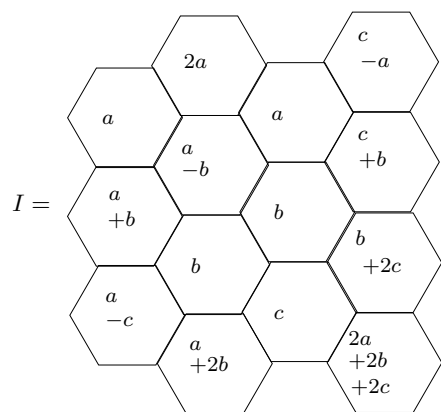
$$\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$$

for some scalars $\alpha_1, \alpha_2, \dots, \alpha_n$.

Notice that if W is a subspace of V , then W is a vector space as well. The commutative, associative, and distributive properties still hold because our scalars are the same and elements of W come from the set V so since these properties hold in V , they hold in W . We say that these properties are *inherited* from V since V is like a parent set to W . The scalar 1 still exists in the scalar set also. This means that we need only show closure under addition and scalar multiplication and that W contains 0. We force W to contain 0. Closure under addition and scalar multiplication hold by choosing appropriate scalars (exercise). Let's look at some examples.

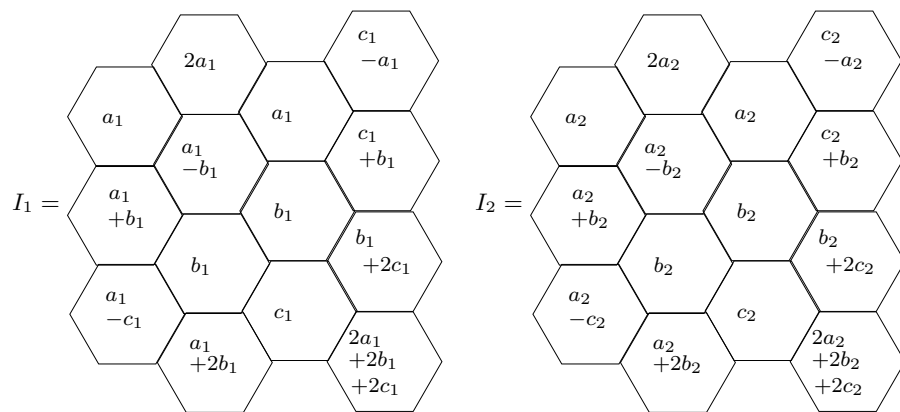
Example 4.6. Consider the set of images

$$V = \{I \mid I \text{ is of the form below and } a, b, c \in \mathbb{R}\}.$$

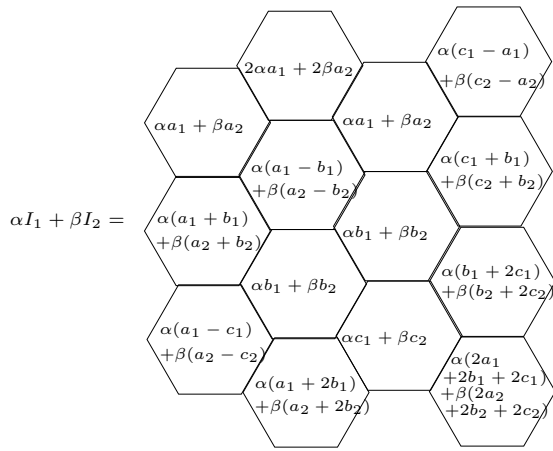


We can show that V is a subspace of images with the same geometric configuration. We showed above that the set of these images is a vector space, so we need only show the two subspace properties.

First, notice that the 0 image is the image with $a = b = c = 0$ and this image is in V . Now, we need to show that linear combinations are still in the set V . Let $\alpha, \beta \in \mathbb{R}$, be scalars and let $I_1, I_2 \in V$, then there are real numbers a_1, b_1, c_1, a_2, b_2 , and c_2 so that



Notice that $\alpha I_1 + \beta I_2$ is also in V since



Notice that performing the operations inside each pixel shows that we can write $\alpha I_1 + \beta I_2$ in the same form as I above. That is, $\alpha I_1 + \beta I_2 \in V$. Thus V is a subspace of images that are laid out in the geometric form above. Notice, this means that V is in itself a vector space.

Example 4.7. Let $V = \{ax^2 + bx + c \mid a + b = 2, a + b + c = 0\}$. Notice that this is a subset of \mathcal{P}_2 . But, notice also that the 0 vector of \mathcal{P}_2 is not in V . We can see this because the 0 vector of \mathcal{P}_2 is $0x^2 + 0x + 0$ and this is not in the set V because $0 + 0 \neq 2$.

Exercises

1. Let $\mathcal{M}_{2 \times 3} = \left\{ \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \mid a, b, c, d, e, f \in \mathbb{R} \right\}$. Show that $\mathcal{M}_{2 \times 3}$ is a vector space when addition and scalar multiplication are defined as in Subsection 2 and scalars are taken from \mathbb{R} .
2. Let $\mathcal{P}_2 = \{ax^2 + bx + c \mid a, b, c \in \mathbb{R}\}$. Show that \mathcal{P}_2 is a vector space with scalars taken from \mathbb{R} and addition and scalar multiplication defined in the standard way for polynomials.
3. Determine whether or not

$$V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \mid x + y + z = 1, 2x + 2y + 2z = 2, \text{ and } -x - y - z = -1 \right\}.$$

is a vector space. Prove your answer.

4. Determine whether or not

$$V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \mid x + y + z = 0, 2x + 2y + 2z = 0, \text{ and } -x - y - z = 0 \right\}.$$

is a vector space. Prove your answer.

5. Make a claim about what types of systems of equations have a solution set that is a vector space.

6. Show that since a subspace, W is closed under linear combinations it is also closed under addition and closed under scalar multiplication, by choosing appropriate scalars.
7. Which of these subsets are subspaces of $\mathcal{M}_{2 \times 2}$? For each that is not, show the condition that fails.

- (a) $\left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \mid a, b \in \mathbb{R} \right\}$
- (b) $\left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \mid a + b = 0 \right\}$
- (c) $\left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \mid a + b = 5 \right\}$
- (d) $\left\{ \begin{pmatrix} a & c \\ 0 & b \end{pmatrix} \mid a + b = 0, c \in \mathbb{R} \right\}$

8. Is this a subspace of \mathcal{P}_2 ?

$$\{a_0 + a_1x + a_2x^2 \mid a_0 + 2a_1 + a_2 = 4\}$$

If it is, prove it. If not, change the set only a little to make it a subspace.

9. Is \mathbb{R}^2 a subspace of \mathbb{R}^3 ? If yes, show it. If no, why not?
10. A manufacturing company uses a process called diffusion welding to adjoin several smaller rods into a single longer rod. The diffusion welding process leaves the final rod heated to various temperatures along the rod with the ends of the rod having the same temperature. Every acm along the rod, a machine records the temperature difference from the temperature at the ends to get an array of temperatures called a *heat state*.
- (a) Plot the heat state given below (let the horizontal axis represent distance from the left end of the rod and the vertical axis represent the temperature difference from the ends).

$$u = (0, 1, 13, 14, 12, 5, -2, -11, -3, 1, 10, 11, 9, 7, 0)$$

- (b) How long is the rod represented by u , the above heat state, if $a = 1cm$?
- (c) Give another example of a heat state for the same rod, sampled in the same locations. (There are many answers you could choose here, this is your choice.)
- (d) Show that the set of all heat states, for this rod, is a vector space. (Hint: You need not show all 10 properties.)
- (e) What do the vectors in this vector space look like? That is, what is the key property (or what are the key properties) that make these vectors stand out?

5 Span

In this section, we discuss linear algebra terminology that will follow us through the rest of this course. In particular, we will define the word span and show some examples. This term gets its own section in these notes for a couple reasons. First, this word is used both as a noun and as a verb in this class. We want to distinguish both. We also want to be clear how to accurately and appropriately use this term.

5.1 Definition of Span

Let us begin with the definition of the *noun* span.

Definition 5.1. (*n.*) Let V be a vector space and let $X = \{v_1, v_2, \dots, v_n\} \subset V$. Then the span of the set X is the set of all linear combinations of the elements of X . That is,

$$\text{span } X = \text{span } \{v_1, v_2, \dots, v_n\} = \{\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n \mid \alpha_1, \alpha_2, \dots, \alpha_n \text{ are scalars}\}.$$

Let us now consider some examples.

Example 5.1. Let us find the span of the two polynomials x and 1 . We get $\text{span } \{x, 1\} = \{ax + b \mid a, b \in \mathbb{R}\} = \mathcal{P}_1$.

Example 5.2. Now, consider the vectors

$$v_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \text{ and } v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^3.$$

We can find the span of $\{v_1, v_2, v_3\}$.

$$\begin{aligned} \text{span } \{v_1, v_2, v_3\} &= \left\{ \alpha \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \mid \alpha, \beta, \gamma \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} -\alpha + \beta \\ 0 \\ \alpha + \gamma \end{pmatrix} \mid \alpha, \beta, \gamma \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} a \\ 0 \\ b \end{pmatrix} \mid a, b \in \mathbb{R} \right\} \\ &= \text{span} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \end{aligned}$$

This example is interesting because it shows two different ways to write the same set as a span. We can also ask whether an element is in a span. We do this in the next example.

Example 5.3. Let

$$v_1 = 3x + 4, v_2 = 2x + 1, v_3 = x^2 + 2, \text{ and } v_4 = x^2.$$

We can determine whether or not $v_1 \in \text{span } \{v_2, v_3, v_4\}$ by deciding whether or not v_1 can be written as a linear combination of v_2, v_3, v_4 . That is, we want to know if there exist scalars α, β , and γ so that

$$v_1 = \alpha v_2 + \beta v_3 + \gamma v_4.$$

If such scalars exist, then

$$3x + 4 = \alpha(2x + 1) + \beta(x^2 + 2) + \gamma(x^2).$$

We can then match up like terms and get the following system of equations:

$$\begin{array}{lcl} (x^2 \text{ term:}) & 0 & = \beta + \gamma \\ (x \text{ term:}) & 3 & = 2\alpha \\ (\text{constant term:}) & 4 & = \alpha + 2\beta \end{array} .$$

Thus $\alpha = \frac{3}{2}, \beta = \frac{5}{4}$, and $\gamma = -\frac{5}{4}$ is a solution to this system. This means that

$$v_1 = \frac{3}{2}v_2 + \frac{5}{4}v_3 - \frac{5}{4}v_4.$$

In other words, $v_1 \in \text{span} \{v_2, v_3, v_4\}$.

Now, before continuing with examples, we should introduce the verb form of the term span.

Definition 5.2. (*v.*) We say that the set of vectors $\{v_1, v_2, \dots, v_n\}$ spans a set X if $X = \text{span} \{v_1, v_2, \dots, v_n\}$. In this case, we call the set $\{v_1, v_2, \dots, v_n\}$ a spanning set of X .

Example 5.4. Notice in Example 5.1, we found that $\{x, 1\}$ spans \mathcal{P}_1 .

Example 5.5. Notice also that $\{x + 1, x - 2, 4\}$ spans \mathcal{P}_∞ . Indeed, we know that if $p \in \text{span} \{x + 1, x - 2, 4\}$ then

$$p = \alpha(x + 1) + \beta(x - 2) + \gamma(4) = (\alpha + \beta)x + \alpha - 2\beta + 4\gamma \in \mathcal{P}_1.$$

Thus, $\text{span} \{x + 1, x - 2, 4\} \subseteq \mathcal{P}_1$. Now if $p \in \mathcal{P}_1$, then $p = ax + b$ for some $a, b \in \mathbb{R}$. We want to show that $p \in \text{span} \{x + 1, x - 2, 4\}$. That is, we want to show that there exist $\alpha, \beta, \gamma \in \mathbb{R}$ so that

$$p = \alpha(x + 1) + \beta(x - 2) + \gamma(4).$$

If such scalars exist, then as before, we can match up like terms to get the system of equations:

$$\begin{array}{l} (x \text{ term:}) \quad a = \alpha + \beta \\ (\text{constant term:}) \quad b = \alpha - 2\beta + 4\gamma \end{array}$$

Thus, if $\alpha = \frac{2a+b}{3}, \beta = \frac{a-b}{3}$, and $\gamma = 0$, then

$$p = \alpha(x + 1) + \beta(x - 2) + \gamma(4).$$

(It should be noted that I noticed there must be infinitely many solutions, I chose $\gamma = 0$ (because 0 is an easy number to work with) and then solved for α and β .) So, for any $p \in \mathcal{P}_1$, we can find such scalars. That means that $\mathcal{P}_1 \subseteq \text{span} \{x + 1, x - 2, 4\}$ and $\{x + 1, x - 2, 4\}$ spans \mathcal{P}_1 .

Example 5.6. Notice that \mathbb{R}^2 is spanned by both $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ and $\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \end{pmatrix} \right\}$.

Indeed, it is clear that

$$\text{span} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \subseteq \mathbb{R}^2 \text{ and } \text{span} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \end{pmatrix} \right\} \subseteq \mathbb{R}^2.$$

Now, if $v = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$, then

$$v = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

So,

$$\mathbb{R}^2 = \text{span} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

Now, we want to show that

$$\begin{pmatrix} a \\ b \end{pmatrix} \in \text{span} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \end{pmatrix} \right\}.$$

So, we find $\alpha, \beta, \gamma \in \mathbb{R}$ so that

$$\begin{pmatrix} a \\ b \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \gamma \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

That is,

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \alpha + \beta - \gamma \\ \alpha + 2\beta + 3\gamma \end{pmatrix}.$$

Thus, we must solve the system of linear equations:

$$\begin{aligned} a &= \alpha + \beta - \gamma \\ b &= \alpha + 2\beta + 3\gamma \end{aligned}$$

Using elimination, we get

$$\begin{array}{rcl} a &= \alpha + \beta - \gamma & \\ b &= \alpha + 2\beta + 3\gamma & \\ \hline & & \xrightarrow{E_2 = -E_1 + E_2} \\ & & b - a = \quad + \beta + 4\gamma \end{array}$$

Again, we see that there are infinitely many α, β, γ that solve this system. Let's pick one. We choose $\gamma = 0$ and so we get $\beta = b - a$ and $\alpha = 2a - b$. Thus

$$\begin{pmatrix} a \\ b \end{pmatrix} = (2a - b) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + (b - a) \begin{pmatrix} 1 \\ 2 \end{pmatrix} + 0 \cdot \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

5.2 Span as a Vector Space

Notice that in each of these examples, we find that the span of a set turned out to be a vector space. It turns out that this is always true.

Theorem 5.1. *Let V be a vector space and let $v_1, v_2, \dots, v_n \in V$. Then $\text{span} \{v_1, v_2, \dots, v_n\}$ is a subspace of V .*

proof We want to show that the two subspace properties hold. That is, we will show that

- $0 \in \text{span} \{v_1, v_2, \dots, v_n\}$.
- $\text{span} \{v_1, v_2, \dots, v_n\}$ is closed under linear combinations.

Notice that $0 = 0 \cdot v_1 + 0 \cdot v_2 + \dots + 0 \cdot v_n \in \text{span} \{v_1, v_2, \dots, v_n\}$. Now, suppose $u, v \in \text{span} \{v_1, v_2, \dots, v_n\}$. We want to show that if α, β are scalars, then $\alpha u + \beta v \in \text{span} \{v_1, v_2, \dots, v_n\}$. Since $u, v \in \text{span} \{v_1, v_2, \dots, v_n\}$, we can find scalars a_1, a_2, \dots, a_n and b_1, b_2, \dots, b_n so that $u = a_1 v_1 + a_2 v_2 + \dots + a_n v_n$ and $v = b_1 v_1 + b_2 v_2 + \dots + b_n v_n$. Now,

$$\begin{aligned} \alpha u + \beta v &= \alpha(a_1 v_1 + a_2 v_2 + \dots + a_n v_n) + \beta(b_1 v_1 + b_2 v_2 + \dots + b_n v_n) \\ &= (\alpha a_1 + \beta b_1) v_1 + (\alpha a_2 + \beta b_2) v_2 + \dots + (\alpha a_n + \beta b_n) v_n \in \text{span} \{v_1, v_2, \dots, v_n\}. \end{aligned}$$

Thus, $\text{span} \{v_1, v_2, \dots, v_n\}$ is closed under linear combinations. \square

The next example relates this back to our first section.

Example 5.7. Let $V = \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0, 3x + 3y + 3z = 0\}$. We can write V as a span. Notice that V is the solution set of the system of equations

$$\begin{array}{rcccc} x & + & y & + & z & = & 0 \\ 3x & + & 3y & + & 3z & = & 0 \end{array}.$$

We see that after elimination, we get the system

$$\begin{array}{rcccc} x & + & y & + & z & = & 0 \\ & & & & 0 & = & 0 \end{array}.$$

Thus y and z can be chosen to be free variables and we get $(x, y, z) = (-y - z, y, z)$. That is,

$$V = \{(-y - z, y, z) \mid y, z \in \mathbb{R}\} = \{y(-1, 1, 0) + z(-1, 0, 1) \mid y, z \in \mathbb{R}\} = \text{span} \{(-1, 1, 0), (-1, 0, 1)\}.$$

Here, we summarize the terminology:

If for every element $v \in V$ we see that $v \in \text{span} \{v_1, v_2, \dots, v_n\}$ this means the same as all of the following statements:

$\{v_1, v_2, \dots, v_n\}$ spans the set V .

$\{v_1, v_2, \dots, v_n\}$ is a spanning set for V .

V is spanned by $\{v_1, v_2, \dots, v_n\}$.

and

There are scalars $\alpha_1, \alpha_2, \dots, \alpha_n$ so that $v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$.

Finally, if $V = \text{span} \{v_1, v_2, \dots, v_n\}$ then all of the above are true and if $w \in \text{span} \{v_1, v_2, \dots, v_n\}$ then $w \in V$.

Exercises

- In Exercise 7 of Section 4, you found that some of the sets were subspaces. For each that was a subspace, write it as a span.
- Decide if the vector lies in the span of the set. If it does, find the linear combination that makes the vector. If it does not, show that no linear combination exists.

(a) $\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ in \mathbb{R}^3 .

(b) $x - x^3, \{x^2, 2x + x^2, x + x^3\}$, in \mathcal{P}_3

(c) $\begin{pmatrix} 0 & 1 \\ 4 & 2 \end{pmatrix}, \left\{ \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 2 & 3 \end{pmatrix} \right\}$, in $\mathcal{M}_{2 \times 2}$

- Which of these sets spans \mathbb{R}^3 ?

(a) $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} \right\}$

- (b) $\left\{ \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$
- (c) $\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix} \right\}$
- (d) $\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix} \right\}$
- (e) $\left\{ \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 6 \\ 0 \\ 2 \end{pmatrix} \right\}$

4. Find a set to span the given subspace

- (a) The xz -plane in \mathbb{R}^3
- (b) $\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid 3x + 2y + z = 0 \right\}$ in \mathbb{R}^3
- (c) $\left\{ \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \mid 2x + y + w = 0 \text{ and } y + 2z = 0 \right\}$ in \mathbb{R}^4
- (d) $\{a_0 + a_1x + a_2x^2 + a_3x^3 \mid a_0 + a_1 = 0 \text{ and } a_2 - a_3 = 0\}$ in \mathcal{P}_3
- (e) The set \mathcal{P}_4 in the space \mathcal{P}_4
- (f) $\mathcal{M}_{2 \times 2}$ in $\mathcal{M}_{2 \times 2}$

5. Answer the following questions with justification.

- (a) If $S \subset T$ are subsets of a vector space is $\text{span}S \subset \text{span}T$? Always? Sometimes? Never?
- (b) If S, T are subsets of a vector space is $\text{span}(S \cup T) = \text{span}S \cup \text{span}T$?
- (c) If S, T are subsets of a vector space is $\text{span}(S \cap T) = \text{span}S \cap \text{span}T$?
- (d) Is the span of the complement equal to the complement of the span?
6. Let's link this with the sections on systems of linear equations. In Exercise 9 Section 1 you found that there were infinitely many solutions to the system. Is any equation in the span of the other two? What about Exercise 13 in Section 1? Is any equation in the span of the other two?
7. Now, use Exercise 6 above to make a similar statement about the rows of the coefficient matrix corresponding to a system of equations.
8. Show (using the allowed operations) that any equation, formed in the elimination process for a system of equations, is in the span of the original equations.

9. Find two different spanning sets (having different number of elements than each other) for each of the following.

- (a) \mathcal{P}_2
(b) $\mathcal{M}_{2 \times 2}$

Which spanning set in each of the above is better? This is really your opinion, but tell me why you think it's better.

6 Linear Independence/Dependence

In the Radiography and Tomography Lab #1, you were asked to write one image using arithmetic operations on three others. You found that this was possible sometimes and not others. It becomes a useful tool to know whether or not this is possible. Let's discuss some terminology first.

6.1 Linear Dependence Definitions

Definition 6.1. Let V be a vector space. We say that the set $\{v_1, v_2, \dots, v_n\} \subset V$, is linearly independent if no element is in the span of the others. That is, if the linear dependence relation

$$\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = 0$$

holds only when

$$\alpha_1 = \alpha_2 = \dots = \alpha_n = 0.$$

We also say that the vectors v_1, v_2, \dots, v_n are linearly independent.

If $\{v_1, v_2, \dots, v_n\}$ is not linearly independent, then we say that it is linearly dependent. Note this means that we can write one element as a linear combination of the others.

Note: The linear dependence relation is definitely true if $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$, but this tells us nothing about the linear dependence of the set. It is important to realize that, to determine the linear dependence, you need to decide whether or not the linear dependence relation is only true when all the scalars are zero. Notice that if the set $\{v_1, v_2\}$ is linearly dependent, then we can find α_1 and α_2 , not both zero, so that $\alpha_1 v_1 + \alpha_2 v_2 = 0$. Suppose that it is α_1 that is not zero. then we can write $v_1 = -\frac{\alpha_2}{\alpha_1} v_2$. That is, v_1 is a scalar multiple of v_2 . Thus, if we are considering the linear dependence a two element set, we need only check whether one can be written as a scalar multiple of the other.

Notice also that the set $\{0, v_1, v_2, \dots, v_n\}$ is always linearly dependent. Indeed,

$$1 \cdot 0 + 0 \cdot v_1 + 0 \cdot v_2 + \dots + 0 \cdot v_n = 0,$$

but one of the scalars (the one on the zero vector) is not zero. Thus, a set containing the zero vector is always linearly dependent. Let us now look at more examples.

Example 6.1. Let us determine whether $\{x + 1, x^2 + 1, x^2 + x + 1\}$ is linearly dependent or independent. We start by setting up the linear dependence relation. We let

$$\alpha(x + 1) + \beta(x^2 + 1) + \gamma(x^2 + x + 1) = 0.$$

Now, we want to find decide whether or not α, β , and γ must all be zero. Matching up like terms in the linear dependence relation leads to the system of equations

$$\begin{aligned} (x^2 \text{ term:}) \quad & 0 = \beta + \gamma \\ (x \text{ term:}) \quad & 0 = \alpha + \gamma. \\ (\text{constant term:}) \quad & 0 = \alpha + \beta + \gamma \end{aligned}$$

Using elimination, we get

$$\begin{array}{rcl} 0 = & \beta + \gamma & 0 = \beta + \gamma \\ 0 = \alpha & + \gamma & \longrightarrow 0 = \alpha + \gamma. \\ 0 = \alpha + \beta + \gamma & & E3 = -e_1 + e_3 \quad 0 = \alpha \end{array}$$

Thus, $\alpha = 0, \beta = 0, \gamma = 0$. This means that $\{x+1, x^2+1, x^2+x+1\}$ is linearly independent.

Example 6.2. Now, let us determine the linear dependence of the set

$$\left\{ \begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} \right\}.$$

Again, we begin by setting up the linear dependence relation. Let

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \alpha \begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix} + \beta \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} + \gamma \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}.$$

We want to find α, β , and γ so that this is true. Matching up entries, we get the following system of equations.

$$\begin{aligned} ((1,1) \text{ entry:}) \quad & 0 = \alpha + \beta + \gamma \\ ((1,2) \text{ entry:}) \quad & 0 = 3\alpha + \beta + 2\gamma \\ ((2,1) \text{ entry:}) \quad & 0 = \alpha + \beta + \gamma. \\ ((2,2) \text{ entry:}) \quad & 0 = \alpha - \beta \end{aligned}$$

We again, use elimination, but this time, let us use a coefficient matrix and reduce it.

$$\begin{array}{l} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 3 & 1 & 2 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & -1 & 0 & 0 \end{array} \right) \xrightarrow[R_3 = -r_1 + r_3, R_4 = -r_1 + r_4]{R_2 = -3r_1 + r_2} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & -2 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -2 & -1 & 0 \end{array} \right) \\ \xrightarrow[R_4 = -r_2 + r_4]{R_2 = \frac{1}{-2}r_2} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \xrightarrow{R_1 = -r_2 + r_1} \left(\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \end{array}$$

Thus, γ can be any real number and $\alpha = -\frac{1}{2}\gamma$ and $\beta = -\frac{1}{2}\gamma$. Thus there are infinitely many possible choices for α, β , and γ . Thus,

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

is a linearly dependent set. Indeed, we found that (choosing $\gamma = 2$)

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

Notice that whenever we are determining whether a set is linearly independent or dependent, we always start with the linear dependence relation and determine whether or not there is only one set of scalars that (when they are all zero) make the linear dependence relation true.

6.2 Basis

In this section, we discuss a way to describe a vector space in terms of a much smaller subset. Our goal is to find a subset that is big enough to describe all of the elements of our vector space, but not so big that there's double up. Let's be more specific. Up until this point, we have seen that a vector space can be written as a span of vectors. That means that we can describe the vector space with all of those vectors. Now, we've also seen that sometimes when we compute the span of a set of vectors, we see that it can be written as a span of a smaller set of vectors. This meant that the vector space can be described with the smaller set of vectors. It turns out this happens when the larger set of vectors is linearly dependent. That is, one or more of the vectors is in the span of the rest of the vectors (there's doubling up, in a sense). Using these, let's define the set that will describe our vector space.

Definition 6.2. *Given a vector space V , we call $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ the basis of V if and only if \mathcal{B} satisfies the following conditions:*

- *span $\mathcal{B} = V$ and*
- *\mathcal{B} is linearly independent.*

Note that the first condition in the definition of a basis gives us that \mathcal{B} is big enough to describe all of V and the second condition says that \mathcal{B} isn't so big that there's doubling up. Let us now look at examples

Example 6.3. *The standard basis for \mathbb{R}^3 is*

$$\mathcal{S} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

First, we show that $\text{span } \mathcal{S} = \mathbb{R}^3$. Notice that $\text{span } \mathcal{S} \subseteq \mathbb{R}^3$. Now, let

$$v = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathbb{R}^3.$$

We see that

$$v = a \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \in \text{span } \mathcal{S}.$$

Thus $\text{span } \mathcal{S} = \mathbb{R}^3$. Now, we show that \mathcal{S} is linearly independent. Let

$$\alpha \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Then

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Thus, $\alpha = \beta = \gamma = 0$. So, \mathcal{S} is linearly independent. We showed both conditions for a basis hold for \mathcal{S} , thus the standard basis is indeed a basis.

Because the standard basis is used often, we introduce notation for each of the vectors. We let e_1, e_2 , and e_3 denote the three vectors in \mathcal{S} , where

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

In general, for \mathbb{R}^n , the standard basis is $\{e_1, e_2, \dots, e_n\}$, where e_i is the $n \times 1$ vector array with zeros in every entry except the i th entry which contains a one.

Example 6.4. *Another basis for \mathbb{R}^3 is*

$$\mathcal{B} = \left\{ \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Notice by Exercise 3b in Section 5, you proved that $\text{span } \mathcal{B} = \mathbb{R}^3$. So, we need only show that \mathcal{B} is linearly independent. Let

$$\alpha \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Then

$$\begin{pmatrix} 2\alpha + \beta \\ \beta \\ \alpha + \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Thus, by matching components, we see that $\alpha = \beta = \gamma = 0$. So, \mathcal{B} is linearly independent and is therefore a basis for \mathbb{R}^3 .

Example 6.5. *The standard basis for \mathcal{P}_2 is $\mathcal{S} = \{x^2, x, 1\}$. Notice in Exercise 1a in Section 6, you showed that \mathcal{S} is linearly independent. Notice also that $\text{span } \mathcal{S} \subseteq \mathcal{P}_2$. So we need to show that $\mathcal{P}_2 \subseteq \text{span } \mathcal{S}$. Let $v = ax^2 + bx + c \in \mathcal{P}_2$. Notice that v is a linear combination of x^2, x , and 1 . Thus $v \in \text{span } \mathcal{S}$. Thus $\text{span } \mathcal{S} = \mathcal{P}_2$. Thus, \mathcal{S} is a basis of \mathcal{P}_2 .*

Example 6.6. *Let $\mathcal{B} = \{1, x + x^2, x^2\}$. In this example, we will show that \mathcal{B} is also a basis for \mathcal{P}_2 . In Exercise 1b in Section 6 you showed that \mathcal{B} is linearly independent. So, we need only show that $\text{span } \mathcal{B} = \mathcal{P}_2$. Notice, we need only show that $\mathcal{P}_2 \subseteq \text{span } \mathcal{B}$ since it is clear that $\text{span } \mathcal{B} \subseteq \mathcal{P}_2$. Let $v = ax^2 + bx + c \in \mathcal{P}_2$. We want to find α, β , and γ so that*

$$\alpha(1) + \beta(x + x^2) + \gamma(x^2) = ax^2 + bx + c.$$

Matching up like terms, we see that $\alpha = c, \beta = b$ and $\beta + \gamma = a$ or $\gamma = a - b$. That is, $v = c(1) + b(x) + (a - b)x^2 \in \text{span } \mathcal{B}$. Thus, \mathcal{B} is a basis for \mathcal{P}_2 .

6.3 Dimension

Notice in the last four examples, we see two examples where the bases of a vector space had the same number of elements. That is, both bases of \mathbb{R}^3 had the same number of elements and both bases for \mathcal{P}_2 had the same number of elements. One should wonder if this is always true. The answer is given in the next theorem.

Theorem 6.1. Let V be a vector space with bases $\mathcal{B}_1 = \{v_1, v_2, \dots, v_n\}$ and $\mathcal{B}_2 = \{u_1, u_2, \dots, u_m\}$. Then the number of elements, n in \mathcal{B}_1 is the same as the number of elements, m in \mathcal{B}_2 .

Proof: Suppose both \mathcal{B}_1 and \mathcal{B}_2 are bases for V having different numbers of elements. We show that this will create an impossible scenario. That is, we will assume that $m \neq n$ and find a reason that there are not two bases of different sizes for V . Suppose $m > n$ (a very similar argument can be made if we assumed $n > m$). Since \mathcal{B}_2 is a subset of V , we know that there exist $\alpha_{i,j}$ for $1 \leq i \leq m$ and $1 \leq j \leq n$ so that

$$\begin{aligned} u_1 &= \alpha_{1,1}v_1 + \alpha_{1,2}v_2 + \dots + \alpha_{1,n}v_n \\ u_2 &= \alpha_{2,1}v_1 + \alpha_{2,2}v_2 + \dots + \alpha_{2,n}v_n \\ &\vdots \\ u_m &= \alpha_{m,1}v_1 + \alpha_{m,2}v_2 + \dots + \alpha_{m,n}v_n. \end{aligned}$$

We want to show that \mathcal{B}_2 cannot be linearly independent (which would be impossible if it is a basis). Let

$$\beta_1 u_1 + \beta_2 u_2 + \dots + \beta_m u_m = 0.$$

We will then find $\beta_1, \beta_2, \dots, \beta_m$. Notice that if we replace u_1, u_2, \dots, u_m with the linear combinations above, we can rearrange to get

$$\begin{aligned} &(\beta_1 \alpha_{1,1} + \beta_2 \alpha_{2,1} + \dots + \beta_m \alpha_{m,1})v_1 \\ &+ (\beta_1 \alpha_{1,2} + \beta_2 \alpha_{2,2} + \dots + \beta_m \alpha_{m,2})v_2 \\ &\quad \vdots \\ &+ (\beta_1 \alpha_{1,n} + \beta_2 \alpha_{2,n} + \dots + \beta_m \alpha_{m,n})v_n = 0. \end{aligned}$$

Since \mathcal{B}_1 is a basis, we get that the coefficients of v_1, v_2, \dots, v_n are all zero. That is

$$\beta_1 \alpha_{1,1} + \beta_2 \alpha_{2,1} + \dots + \beta_m \alpha_{m,1} = 0 \tag{1}$$

$$\beta_1 \alpha_{1,2} + \beta_2 \alpha_{2,2} + \dots + \beta_m \alpha_{m,2} = 0 \tag{2}$$

$$\quad \quad \quad \vdots \tag{3}$$

$$\beta_1 \alpha_{1,n} + \beta_2 \alpha_{2,n} + \dots + \beta_m \alpha_{m,n} = 0. \tag{4}$$

We know that this system has a solution because it is homogeneous. But, because there are more scalars $\beta_1, \beta_2, \dots, \beta_m$ (that we are solving for) than there are equations, this system must have infinitely many solutions. This means that \mathcal{B}_2 cannot be linearly independent and so it cannot be a basis.

Thus, the only way both \mathcal{B}_1 and \mathcal{B}_2 can be bases of the same vector space is if they have the same number of elements. \square

Because the number of elements in a basis is unique to the vector space, we can give it a name.

Definition 6.3. Given a vector space V whose basis \mathcal{B} has n elements, we say that the dimension of V is n , the number of elements in \mathcal{B} ($\dim V = n$.) We also say that V is n -dimensional.

Notice that when proving Theorem 6.1, we saw that an n -dimensional vector space V has no linearly independent set with more than n elements. That is, if we have a set $\{u_1, u_2, \dots, u_k\} \subset V$ and $k > n$, then we automatically know that the set is linearly dependent. This gives us another tool to make a quick check of linear dependence and may save us time.

In the next few examples, we illustrate how we find the dimension of various vector spaces.

Example 6.7. *Let*

$$V = \left\{ \left(\begin{array}{c} x \\ y \\ z \end{array} \right) \mid x + y + z = 0, 2x + y - 4z = 0, 3x + 2y - 3z = 0 \right\}.$$

Notice that V is the solution set of a homogeneous system of equations. So, we know that V is a subspace (and therefore a vector space) of \mathbb{R}^3 . Notice also V has dimension 1. We show this below.

First, we need to rewrite V as a span. We can reduce the system of equations

$$\begin{array}{rcl} x + y + z = 0 \\ 2x + y - 4z = 0 \\ 3x + 2y - 3z = 0 \end{array} \xrightarrow{\text{use matrix}} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 2 & 1 & -4 & 0 \\ 3 & 2 & -3 & 0 \end{array} \right) \begin{array}{l} R_2 = -2r_1 + r_2 \\ R_3 = -3r_1 + r_3 \end{array} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & -1 & -6 & 0 \\ 0 & -1 & -6 & 0 \end{array} \right) \\ \xrightarrow{R_1 = r_2 + r_1, R_3 = -r_2 + r_3, R_2 = -r_2} \left(\begin{array}{ccc|c} 1 & 0 & -5 & 0 \\ 0 & 1 & 6 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \longrightarrow \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 5z \\ -6z \\ z \end{pmatrix},$$

where z can be any real number. This means we can rewrite V as below

$$V = \left\{ \left(\begin{array}{c} 5z \\ -6z \\ z \end{array} \right) \mid z \in \mathbb{R} \right\} = \text{span} \left\{ \left(\begin{array}{c} 5 \\ -6 \\ 1 \end{array} \right) \right\}.$$

Notice that the set

$$\mathcal{B} = \left\{ \left(\begin{array}{c} 5 \\ -6 \\ 1 \end{array} \right) \right\}$$

is linearly independent and spans V . Thus it is a basis for V . This tells us that V has dimension 1.

Example 6.8. *Let $V = \{ax^2 + bx + c \mid a + b - 2c = 0\}$. We know that V is a subspace of \mathcal{P}_2 . Indeed, $0x^2 + 0x + 0 \in V$ and if $v_1 = a_1x^2 + b_1x + c_1$ and $v_2 = a_2x^2 + b_2x + c_2$ are vectors in V and α, β be scalars, then*

$$a_1 + b_1 - 2c_1 = 0 \quad \text{and} \quad a_2 + b_2 - 2c_2 = 0.$$

Now,

$$\alpha v_1 + \beta v_2 = (\alpha a_1 + \beta a_2)x^2 + (\alpha b_1 + \beta b_2)x + \alpha c_1 + \beta c_2.$$

Also, we see that

$$\alpha a_1 + \beta a_2 + \alpha b_1 + \beta b_2 + 2\alpha c_1 + 2\beta c_2 = \alpha(a_1 + b_1 - 2c_1) + \beta(a_2 + b_2 - 2c_2) = 0 + 0 = 0.$$

Thus, $\alpha v_1 + \beta v_2 \in V$. Now, below, we show that V is 2-dimensional. Indeed, we can rewrite V :

$$V = \{(2c - b)x^2 + bx + c \mid b, c \in \mathbb{R}\} = \{(-x^2 + x)b + (2x^2 + 1)c \mid b, c \in \mathbb{R}\} = \text{span} \{-x^2 + x, 2x^2 + 1\}.$$

Now, we can see that the elements of the set $\mathcal{B} = \{-x^2 + x, 2x^2 + 1\}$ are not scalar multiples of one another so therefore \mathcal{B} is linearly independent. Thus, \mathcal{B} is a basis for V and so V is 2-dimensional.

What we see is that, in order to find the dimension of a vector space, we need to find a basis and count the elements in the basis.

Exercises

- Determine whether the following sets are linearly independent.

(a) $\{1, x, x^2\}$

(b) $\{1, x + x^2, x^2\}$

(c) $\{1, 1 - x, 1 + x, 1 + x^2\}$

- Determine whether each of the following sets is linearly independent.

(a) $\left\{ \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$

(b) $\left\{ \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 2 \\ 1 & 0 \end{pmatrix} \right\}$

- Suppose $\{v_1, v_2, v_3, v_4\}$ are linearly independent. Determine if the following sets are linearly independent. Justify your answer. If not, remove only enough vectors to make the set independent.

(a) $\{v_1, v_2\}$

(b) $\{v_1, v_2, v_3, v_4, v_1 - 2v_3\}$

(c) $\{v_1 + v_3, v_2 + v_4, v_3, v_4\}$

(d) $\{v_1 - 2v_2, v_2, v_3 - v_4 - v_2, v_4\}$

- Determine which of the following sets is a basis for \mathcal{P}_2 .

(a) $\{1, x + x^2, x^2\}$

(b) $\{x^2 - 1, 1 + x, 1 - x\}$

(c) $\{1, 1 - x, 1 + x, 1 + x^2\}$

- Suppose $\{v_1, v_2, v_3, v_4\}$ are linearly independent. Determine which of the following sets is a basis for $\text{span}\{v_1, v_2, v_3, v_4\}$. Justify your answer.

(a) $\{v_1, v_2\}$

(b) $\{v_1, v_2, v_3, v_4, v_1 - 2v_3\}$

- (c) $\{v_1 + v_3, v_2 + v_4, v_3, v_4\}$
 (d) $\{v_1 - 2v_2, v_2, v_3 - v_4 - v_2, v_4\}$

6. For each of the following decide whether or not \mathcal{B} is a basis for the vector space V .

- (a) $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}, V = \mathbb{R}^3$
 (b) $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}, V = \mathbb{R}^2$
 (c) $\mathcal{B} = \left\{ \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right\}, V = \mathcal{M}_{2 \times 2}$
 (d) $\mathcal{B} = \{x^2, x^2 + x, x^2 + x + 2\}, V = \mathcal{P}_2$

7. For each of the vector spaces below, find basis \mathcal{B} that is not the standard basis. State the dimension of the vector space.

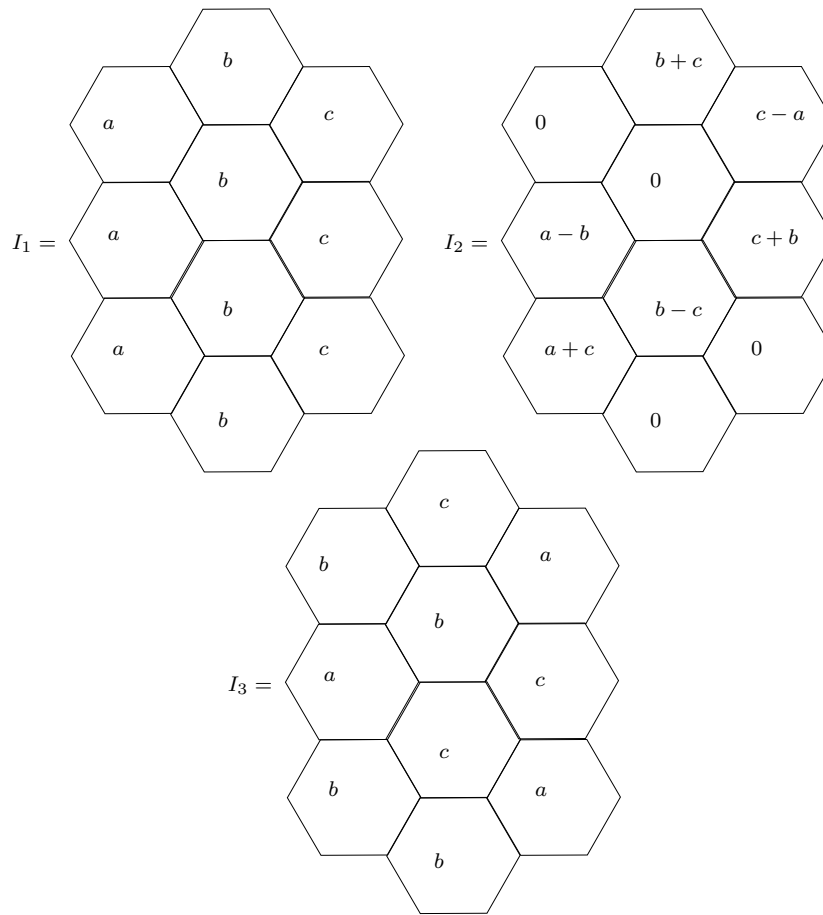
- (a) $\left\{ \begin{pmatrix} a & c \\ 3d & b \end{pmatrix} \mid a + b + c - 2d = 0, a + 3b - 4c + d = 0, a - d + b = c \right\}$
 (b) $\{cx^2 + 3bx - 4a \mid a - b - 2c = 0\}$
 (c) $\mathcal{M}_{3 \times 2}$
 (d) \mathcal{P}_3

- (e) $\text{span} \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$

8. Given the set $\mathcal{B} = \{u, v, w\}$. Show that if \mathcal{B} is a basis, then so is

$$\mathcal{B}' = \{u + 2v, u - w, v + w\}.$$

9. Using #8, make a general statement about how to get a basis from another basis. Be careful to use accurate linear algebra language.
 10. Determine whether $\mathcal{B} = \{I_1, I_2, I_3\}$, where the I_n are given below, is a basis for the vector space of images with the same orientation as each of the I_n below.



11. What is the dimension of the vector space of heat signatures given in Exercise 10 in Section 4?
12. Show that given a homogeneous system of linear equations with 3 variables, that the system has only the trivial solution whenever the corresponding matrix when written as a matrix equation $Ax = b$ has linearly independent rows.

7 Coordinate Spaces

7.1 Coordinates in the Standard Basis

In our typical 3D space, we talk about vectors that look like $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$. And we see that the coordinates are x , y , and z respectively. We talk about this like it just makes sense, but we should really spell out what this means. We mean that the vector points from the origin to a point that is x units horizontally, y units vertically, and z units up from the origin (as in Figure 2. The truth is that when we say that, we are assuming that we are working with the standard basis. (This makes sense because, it is the basis that we usually think about, x -axis perpendicular to the y -axis, forming a horizontal plane and the z axis perpendicular to this plane.)

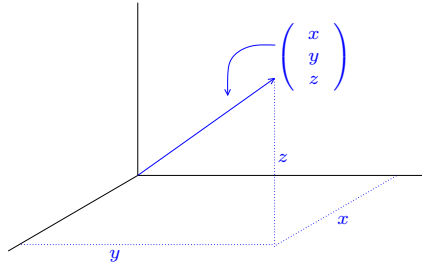


Figure 2:

So, we can write

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

Notice that the coordinates are the scalars in this linear combination. That's what we mean by coordinates.

7.2 Coordinates in Other Bases

Now, let us consider a different basis,

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

The coordinates of the vector $v = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ in the standard basis are 1, 2, and 3 respectively,

but in this new basis, they are found by finding the scalars α_1, α_2 , and α_3 so that

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \alpha_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Going through the motions to solve for α_1, α_2 , and α_3 , we find that $\alpha_1 = 2, \alpha_2 = -1$, and $\alpha_3 = 2$. (Be sure to check this yourself.) So, we can represent the vector v in coordinates according to the basis \mathcal{B} as

$$[v]_{\mathcal{B}} = \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix}.$$

Notice that we indicate that we are looking for coordinates in terms of the basis \mathcal{B} by using the notation, $[v]_{\mathcal{B}}$.

This means if we are given a vector space V and bases $\mathcal{B}_1 = \{v_1, v_2, \dots, v_n\}$ and $\mathcal{B}_2 = \{u_1, u_2, \dots, u_n\}$, then if $w = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$, we have

$$[w]_{\mathcal{B}_1} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

But, if $w = \beta_1 u_1 + \beta_2 u_2 + \dots + \beta_n u_n$, we have

$$[w]_{\mathcal{B}_2} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}.$$

It is really important, then, to always note the basis in which you are describing your vectors. Notice also that coordinate vectors look like vectors in \mathbb{R}^n for some n . Let us look at some more examples.

Example 7.1. Let $V = \{ax^2 + (b)x + (c) \in \mathcal{P}_2 \mid a + b - 2c = 0\}$. We saw in Example 6.8 that a basis for V is $\mathcal{B} = \{-x^2 + x, 2x^2 + 1\}$. We also know that $v = 3x^2 - 3x \in V$. This means that we can write v as a coordinate vector. We want to find α_1, α_2 so that

$$v = \alpha_1(-x^2 + x) + \alpha_2(2x^2 + 1).$$

Looking at this in the right way makes it clear that

$$v = -3(-x^2 + x) + 0(2x^2 + 1).$$

Thus

$$[v]_{\mathcal{B}} = \begin{pmatrix} -3 \\ 0 \end{pmatrix}.$$

Now, let

$$[w]_{\mathcal{B}} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

Then we can find w , using our understanding of the coordinate vector and basis, as follows.

$$w = 2(-x^2 + x) - 1(2x^2 + 1) = -4x^2 + 2x - 1.$$

We can check that $w \in V$. Indeed, $-4 + 2 + 2(-1) = 0$, thus $w \in V$.

Let's now consider representations of a vector when we view the vector span in terms of two different bases.

Example 7.2. Let

$$V = \left\{ \begin{pmatrix} a & a+b \\ a-b & b \end{pmatrix} \mid a, b \in \mathbb{R} \right\}.$$

Then we can find two bases for V . First, we write V as a span.

$$V = \left\{ a \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} + b \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} \mid a, b \in \mathbb{R} \right\} = \text{span} \left\{ \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} \right\}.$$

Notice that the set

$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} \right\}$$

is a basis for V . Note also that

$$\mathcal{B}_2 = \left\{ \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix} \right\}$$

is also a basis for V . (Check to see that you actually believe this.) Now, consider the vector w so that

$$[w]_{\mathcal{B}_1} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Then, we know that

$$w = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} + 2 \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ -1 & 2 \end{pmatrix}.$$

If we want to write w in terms of the basis \mathcal{B}_2 , we need to find scalars α and β so that

$$\text{and } [w]_{\mathcal{B}_2} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

That is, we want α and β so that

$$w = \begin{pmatrix} 1 & 3 \\ -1 & 2 \end{pmatrix} = \alpha \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} + \beta \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}.$$

By inspecting the components, we see that $\alpha = \frac{3}{2}$ and $\beta = -\frac{1}{2}$. Thus,

$$[w]_{\mathcal{B}_2} = \frac{1}{2} \begin{pmatrix} 3 \\ -1 \end{pmatrix}.$$

Example 7.3. Let $V = \{ax^2 + (b-a)x + (a+b) \mid a, b \in \mathbb{R}\}$. Then

$$V = \{a(x^2 - x + 1) + b(x + 1) \mid a, b \in \mathbb{R}\} = \text{span}\{x^2 - x + 1, x + 1\}.$$

So a basis for V is

$$\mathcal{B} = \{v_1 = x^2 - x + 1, v_2 = x + 1\}.$$

This means that $\dim V = 2$ and so vectors in V can be represented by vectors in \mathbb{R}^2 . Notice that $v = 3x^2 + x + 7 \in V$. You actually should be able to check this. We can write v in terms of the basis of V as $v = 3v_1 + 4v_2$. We can check this as follows $3v_1 + 4v_2 = 3(x^2 - x + 1) + 4(x + 1) = 3x^2 + x + 7$. Thus, the coordinate vector for v is $[v]_{\mathcal{B}} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$.

Example 7.4. Let $W = \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \alpha + \beta + \gamma \end{pmatrix} \mid \alpha, \beta, \gamma \in \mathbb{R} \right\}$. A basis, $\tilde{\mathcal{B}}$, for W is

$$\tilde{\mathcal{B}} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \right\}.$$

We can determine if $w = \begin{pmatrix} 3 & 4 \\ -1 & 6 \end{pmatrix} \in W$. If it is, we can find the coordinate vector, $[w]_{\tilde{\mathcal{B}}}$ in \mathbb{R}^3 .

First, we determine if $w \in W$. Let v_1, v_2 , and v_3 be the above basis elements. Then we write $w = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3$ (and check to see if it is possible to find α_1, α_2 , and α_3 .)

$$\begin{pmatrix} 3 & 4 \\ -1 & 6 \end{pmatrix} = \alpha_1 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}.$$

This leads to the equations:

$$\begin{aligned}3 &= \alpha_1 \\4 &= \alpha_2 \\-1 &= \alpha_3 \\6 &= \alpha_1 + \alpha_2 + \alpha_3\end{aligned}$$

These all are true if the first three equations are true. So, we know that $w \in W$ and in this process, we found that $[w]_{\mathcal{B}} = \begin{pmatrix} 3 \\ 4 \\ -1 \end{pmatrix}$.

Using other coordinate systems can be very useful. Sometimes the coordinates themselves are just nicer numbers. We'll see more use of this concept later.

Exercises

1. Using the vector space and corresponding basis in Example 7.3, find the coordinate vector for $\tilde{v} = 5x^2 - 7x + 3$.
2. Let $X = \{(x, y, x - y, x + y) | x, y \in \mathbb{R}\}$. Find a basis, \mathcal{B} for X . Determine if $v = (2, 3, 1, 5) \in X$. If so, find $[v]_{\mathcal{B}}$.
3. Given V with basis $\mathcal{B} = \{2x + 1, x^2, 2x^3\}$, what element of $w \in V$ is

$$[w]_{\mathcal{B}} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}.$$

4. Given V so that $V = \text{span } \mathcal{B}$, where

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \right\}.$$

Verify that \mathcal{B} is indeed a basis for V . Find the element $v \in V$ so that

$$[v]_{\mathcal{B}} = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}.$$

5. Given $V = \text{span } \{v_1, v_2, v_3, v_4\}$. Suppose also that $\mathcal{B} = \{v_1, v_2, v_3, v_4\}$ is linearly independent. Let $u = v_1 + 2v_2 - 3v_3 - v_4$. Write the coordinate vector $[u]_{\mathcal{B}}$.
6. Using the vector space and vector u in Exercise 5, with basis

$$\mathcal{B}_2 = \{v_1 - v_2, v_1 + v_2 + 3v_3 - v_4, v_2, v_3\},$$

find $[u]_{\mathcal{B}_2}$.

7. Recall Exercise 10 in Section 4.2. Let V be a vector space of heat states with basis

$$\mathcal{B} = \left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 2 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -3 \\ 0 \end{pmatrix} \right\}.$$

Suppose that u is a heat state in V and that

$$u = \begin{pmatrix} 0 \\ 1 \\ -2 \\ -3 \\ -1 \\ 0 \end{pmatrix}.$$

(Be sure you can write a proof to show that \mathcal{B} is indeed a basis.) What is the dimension V ? Write the coordinate vector $[u]_{\mathcal{B}}$.

8 Transmission Radiography and Tomography A Simplified Overview

This material provides a brief overview of radiographic principles prerequisite to Lab #2 of the Radiography and Tomography Linear Algebra Modules. The goal is to develop the basic discrete radiographic operator for axial tomography of the human body. To accomplish this goal, it is **not** necessary to completely understand the details of the physics and engineering involved here. We wish to arrive at a mathematical formulation descriptive of the radiographic process and establish a standard scenario description with notation. Read through these notes and answer the questions at the end.

8.1 What is Radiography?

Transmission radiography and tomography are familiar and common processes in today's world, especially in medicine and non-destructive testing in industry. Some examples include

- Single-view X-ray radiography is used routinely to view inside the human body; for example, bone fracture assessment, mammography, and angiographic procedures.
- Multiple-view X-ray radiography is realized in computerized axial tomography (CAT) scans used to provide 3D images of body tissues.
- Neutron and X-ray imaging is used in industry to quantify manufactured part assemblies or defects which cannot be visually inspected.

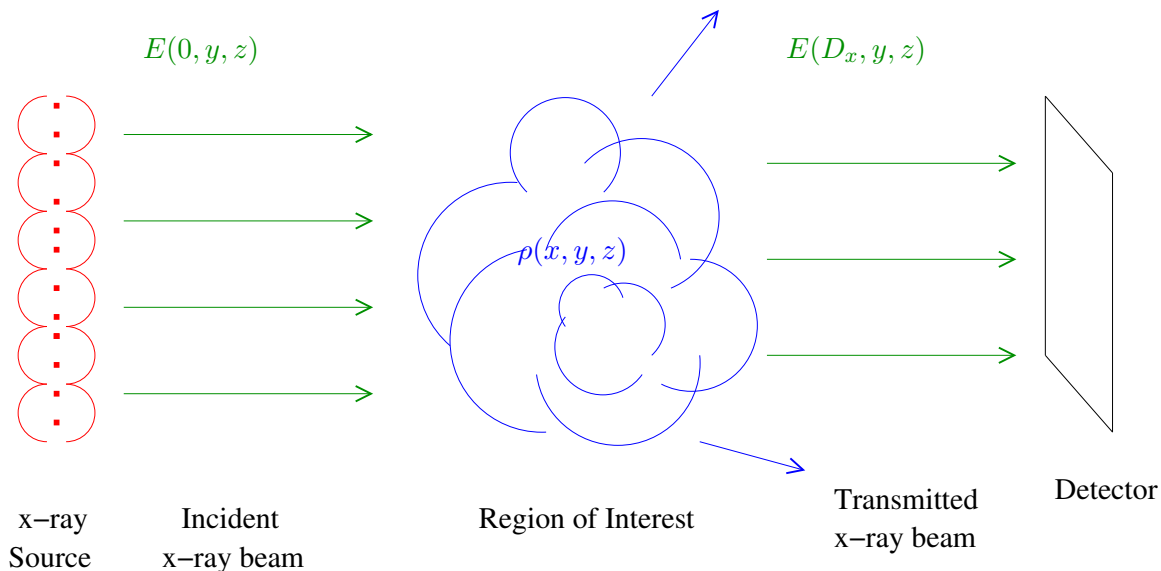


Figure 3: Typical radiographic experiment.

8.1.1 Transmission Radiography

is the process of measuring and recording changes in a high-energy particle beam (X-rays, protons, neutrons, etc.) resulting from passage through an object of interest.

8.1.2 Tomography

is the process of inferring properties of an unknown object by interpreting radiographs of the object.

X-rays, just like visible light, are photons or electromagnetic radiation, but at much higher energies and outside of the range of our vision. Because of the wavelength of typical X-rays (on the order of a nanometer), they readily interact with objects of similar size such as individual molecules or atoms. This property makes them particularly useful in transmission imaging. Figure 3 is a cartoon of a typical x-ray radiographic experiment or procedure. An x-ray beam is produced with known energy and geometric characteristics. The beam is aimed at a region of interest. The photons interact with matter in the region of interest, changing the intensity, energy and geometry of the beam. A detector measures the pattern (and possibly the distribution) of incident energy. The detection data, when compared to the incident beam characteristics, contains the known signature of the region of interest. We consider the mathematics and some of the physics involved in each step with the goal of modeling a radiographic transformation appropriate for mixed soft and hard tissue axial tomography of the human body.

8.2 The Incident X-ray Beam

We begin with an x-ray beam in which the x-ray photons all travel parallel to each other in the beam direction, which we take to be the positive x -direction. Additionally we assume that the beam is of short time duration, the photons being clustered in a short pulse instead of being continuously produced. A beam with these geometric characteristics is usually not directly achievable through typical x-ray sources (see supplementary material for some

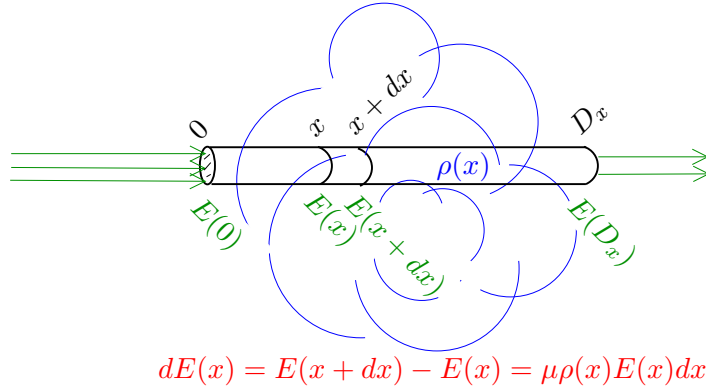


Figure 4: X-ray beam attenuation computation for a beam path of fixed y and z .

discussion on x-ray sources). However, such a beam can be approximated readily through so-called collimation techniques which physically limit the incident x-rays to a subset that compose a planar (neither convergent nor divergent) beam.

While not entirely necessary for the present formulation, we consider a monochromatic x-ray source. This means that every x-ray photon produced by the source has exactly the same “color.” The term “monochromatic” comes from the visible light analog in which, for example, a laser pointer may produce photons of only one color, red. The energy of a single photon is proportional to its frequency ν , or inversely proportional to its wavelength λ . The frequency, or wavelength, determine the color, in exact analogy with visible light. In particular, the energy of a single photon is $h\nu$ where the constant of proportionality h is known as Planck’s constant.

The intensity (or brightness), $E(x, y, z)$, of a beam is proportional to the photon density. The intensity of the beam just as it enters the region of interest at $x = 0$ is assumed to be the same as the intensity at the source. We write both as $E(0, y, z)$. It is assumed that this quantity is well known or independently measurable.

8.3 X-Ray Beam Attenuation

As the beam traverses the region of interest, from $x = 0$ to $x = D_x$ (see Figure 4), the intensity changes as a result of interactions with matter. In particular, a transmitted (resultant) intensity $E(D_x, y, z)$ exits the far side of the region of interest. We will see that under reasonable assumptions this transmitted intensity is also planar but is reduced in magnitude. This process of intensity reduction is called attenuation. It is our goal in this section to model the attenuation process.

X-ray attenuation is a complex process that is a combination of several physical mechanisms (outlined in the supplementary material) describing both scattering and absorption of x-rays. We will consider only Compton scattering which is the dominant process for typical medical x-ray radiography. In this case, attenuation is almost entirely a function of photon energy and material mass density. As we are considering monochromatic (monoenergetic) photons, attenuation is modeled as a function of mass density only.

Consider the beam of initial intensity $E(0, y, z)$ passing through the region of interest, at fixed y and z . The relative intensity change in an infinitesimal distance from x to $x + dx$ is proportional to the mass density $\rho(x, y, z)$ and is given by

$$dE(x, y, z) = E(x + dx, y, z) - E(x, y, z) = -\mu\rho(x, y, z)E(x, y, z)dx$$

where μ is a factor that is nearly constant for many materials of interest. We also assume that any scattered photons exit the region of interest without further interaction. This is the so-called single-scatter approximation which dictates that the intensity remains planar for all x .

Integrating over the path from $x = 0$ where the initial beam intensity is $E(0, y, z)$ to $x = D_x$ where the beam intensity is $E(D_x, y, z)$ yields

$$\begin{aligned}\frac{dE(x, y, z)}{E(x, y, z)} &= -\mu\rho(x, y, z)dx \\ \int_{E(0, y, z)}^{E(D_x, y, z)} \frac{dE(x, y, z)}{E(x, y, z)} &= -\mu \int_0^{D_x} \rho(x, y, z)dx \\ \ln E(D_x, y, z) - \ln E(0, y, z) &= -\mu \int_0^{D_x} \rho(x, y, z)dx \\ E(D_x, y, z) &= E(0, y, z)e^{-\mu \int_0^{D_x} \rho(x, y, z)dx}.\end{aligned}$$

This expression shows us how the initial intensity is reduced, because photons have been scattered out of the beam. The relative reduction depends on the density (or mass) distribution in the region of interest.

8.4 Radiographic Energy Detection

The transmitted intensity $E(D_x, y, z)$ continues to travel on to a detector (e.g. film) which records the total detected energy in each of m detector bins. The detected energy in any bin is the intensity integrated over the bin cross-sectional area. Let p_k be the number of x-ray photons collected at detector bin k . p_k is then the collected intensity integrated over the bin area and divided by the photon energy.

$$p_k = \frac{1}{h\nu} \iint_{(bin\ k)} E(0, y, z) \left(e^{-\mu \int_0^{D_x} \rho(x, y, z)dx} \right) dydz.$$

Let the bin cross sectional area, σ , be small enough so that both the contributions of the density and intensity to the bin area integration are approximately a function of x only. Then

$$p_k = \frac{\sigma E(0, y_k, z_k)}{h\nu} e^{-\mu \int_0^{D_x} \rho(x, y_k, z_k)dx},$$

where y_k and z_k locate the center of bin k . Let p_k^0 be the number of x-ray photons initially aimed at bin k , $p_k^0 = \sigma E(0, x, y)/h\nu$. Due to attenuation, $p_k \leq p_k^0$ for each bin.

$$p_k = p_k^0 e^{-\mu \int_0^{D_x} \rho(x, y_k, z_k)dx}.$$

Equivalently, we can write (multiply the exponent argument by σ/σ):

$$p_k = p_k^0 e^{-\frac{\mu}{\sigma} \int_0^{D_x} \sigma \rho(x, y_k, z_k)dx}.$$

The remaining integral is the total mass in the region of interest that the x-ray beam passes through to get to bin k . We will call this mass s_k . Now we have

$$p_k = p_k^0 e^{-s_k/\alpha},$$

where $\alpha = \sigma/\mu$. This expression tells us that the number of photons in the part of the beam directed at bin k is reduced by a factor that is exponential in the total mass encountered by the photons.

Finally, we note that the detector bins correspond precisely to pixels in a radiographic image.

8.5 The Radiographic Transformation Operator

We consider a region of interest subdivided into N cubic voxels (three-dimensional pixels). Let x_j be the mass in object voxel j and T_{kj} the fraction of voxel j in beam path k (see Figure 5). Then the mass along beam path k is

$$s_k = \sum_{j=1}^N T_{kj} x_j,$$

and the expected photon count at radiograph pixel k , p_k , is given by

$$p_k = p_k^0 e^{-\frac{1}{\alpha} \sum_{j=1}^N T_{kj} x_j},$$

or equivalently,

$$b_k \equiv \left(-\alpha \ln \frac{p_k}{p_k^0} \right) = \sum_{j=1}^N T_{kj} x_j.$$

The new quantities b_k represent a variable change that allows us to formulate the matrix expression for the radiographic transformation

$$b = Tx.$$

This expression tells us that given a voxelized object mass distribution image $x \in \mathbb{R}^N$, the expected radiographic data (mass projection) is image $b \in \mathbb{R}^m$, with the two connected through radiographic transformation $T \in \mathcal{M}_{m \times N}(\mathbb{R})$. The mass projection b and actual photon counts p and p^0 are related as given above. It is important to note that b_k is defined only for $p_k > 0$. Thus, this formulation is only valid for radiographic scenarios in which every radiograph detector pixel records at least one hit. This is always the case for medical applications which require high contrast and high signal-to-noise ratio data.

8.6 Multiple Views and Axial Tomography

Thus far, we have a model that can be used to compute a single radiograph of the region of interest. In many applications it is beneficial to obtain several or many different views of this region. In some industrial applications, the region of interest can be rotated within the radiographic apparatus, with radiographs obtained at various rotation angles. In medical applications the radiographic apparatus is rotated about the region of interest (including the subject!). In this latter case, the voxelization remains fixed and the coordinate system

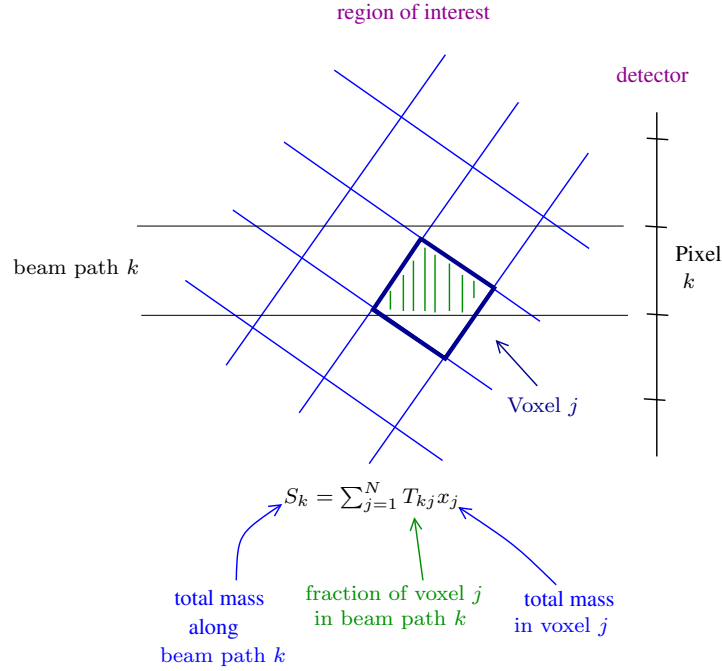


Figure 5: Object space and radiograph space discretization.

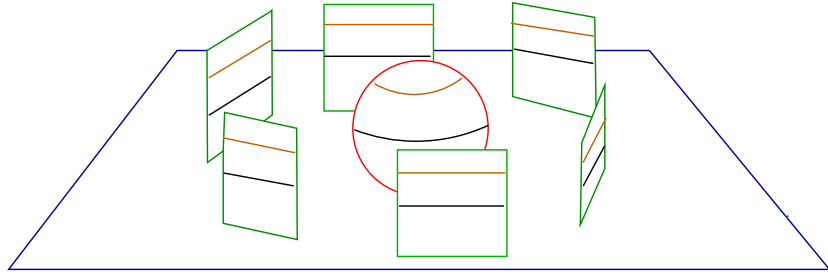


Figure 6: Example axial radiography scenario with six equally spaced views. Horizontal slices of the object project to horizontal rows in the radiographs.

rotates. For each of a view angles the new m pixel locations require calculation of new mass projections T_{kj} . The full multiple-view operator contains mass projections onto all $M = a \cdot m$ pixel locations. Thus, for multiple-view radiography: x is still a vector in \mathbb{R}^N , but b is a vector in \mathbb{R}^M and T is a matrix operator in $\mathcal{M}_{M \times N}(\mathbb{R})$.

Finally, we make the distinction between the general scenario and axial tomography (CAT scans). In principle, we could obtain radiographs of the region of interest from any direction (above, below, left, right, front, back, etc.). However, in axial tomography the physical limitations of the apparatus and subject placement dictate that views from some directions are not practical. The simplest scenario is to obtain multiple views by rotating the apparatus about a fixed direction perpendicular to the beam direction. This is why CAT machines have a donut or tube shaped appearance within which the apparatus is rotated. The central table allows the subject to rest along the rotation axis and the beam can pass through the subject along trajectories.

This axial setup also simplifies the projection operator. If we consider the ℓ^{th} slice of the region of interest, described by an $n \times n$ array of N voxels, the mass projections of this slice will only occur in the ℓ^{th} row of pixels in each radiographic view see Figure 6. As

a result, 3D reconstructions can be obtained by a series of independent 2D reconstructed slices. For example, the brown slice of the spherical object (represented in \mathbb{R}^N) is related to the collection of brown rows of the radiographs (represented in \mathbb{R}^M) through the projection operator $T \in \mathcal{M}_{M \times N}(\mathbb{R})$. The black slice and black rows are related through the *same* projection operator.

8.7 Model Summary

The list below gathers the various mathematical quantities of interest.

- N is the number of object voxels.
- M is the number of radiograph pixels.
- $x \in \mathbb{R}^N$ is the material mass in each object voxel.
- $b \in \mathbb{R}^M$ is the mass projection onto each radiograph pixel.
- $p \in \mathbb{R}^M$ is the photon count recorded at each radiograph pixel.
- $T \in \mathcal{M}_{N \times M}(\mathbb{R})$ is voxel volume projection operator. T_{ij} is the fractional volume of voxel j which projects orthogonally onto pixel i .
- p_k^0 is the incident photon count per radiograph pixel.
- $b = -\alpha \ln \frac{p}{p^0}$.
- $b = Tx$ is the (mass projection) radiographic transformation.

The description of images (objects and radiographs) as vectors in \mathbb{R}^N and \mathbb{R}^M is computationally useful and more familiar than a vector spaces of images. One should keep in mind that this is a particular representation for images which is useful as a tool but is not geometrically descriptive. The price we pay for this convenience is that we no longer have the geometry of the radiographic setup (pixelization and voxelization) encoded in the representation.

A vector in \mathbb{R}^3 , say $(1, 2, 5)$, is a point in a three-dimensional space with coordinates described relative to three orthogonal axes. We can actually locate this point and plot it. An image represented in \mathbb{R}^3 , say $(1, 2, 5)$, is *not* a point in this space. Without further information about the vector space of which it is a member, we cannot draw this image. The use of \mathbb{R}^3 allows us to perform scalar multiplication and vector addition on images *because these operations are equivalently defined on \mathbb{R}^3* .

8.8 Model Assumptions

The radiography model we have constructed is based on a number of approximations and assumptions which we list here. This list is not comprehensive, but it does gather the most important concepts.

- **Monochromaticity.** Laboratory sources generate x-rays with a wide range of energies as a continuous spectrum. We say that such x-ray beams are polychromatic. One method of approximating a monochromatic beam is to precondition the beam by having the x-rays pass through a uniform material prior to reaching the region of interest. This process preferentially attenuates the lower energy photons, leaving only the highest energy photons. This process is known as beam-hardening. The result is a polychromatic beam with a narrower range of energies. We can consider the beam to be approximately monochromatic, especially if the attenuation coefficient(s) of the material, μ , is not a strong function of photon energy.
- **Geometric Beam Characteristics.** Laboratory sources do not naturally generate planar x-ray beams. It is more characteristic to have an approximate point source with an intensity pattern that is strongly directionally dependent. Approximate planar beams with relatively uniform intensity $E(0, y, x)$ can be achieved by selective beam shielding and separation of source and region of interest. In practice, it is common to use the known point source or line source characteristics instead of assuming a planar beam. The model described here is unchanged except for the computation of T itself.
- **Secondary Radiation.** Our model uses a single-scatter approximation in which if a photon undergoes a Compton scatter, it is removed from the analysis. In fact, x-rays can experience multiple scatter events as they traverse the region of interest or other incidental matter (such as the supporting machinery). The problematic photons are those that scatter one or more times *and* reach the detector. This important secondary effect is often approximated by more advanced models.
- **Energy-Dependent Attenuation.** The attenuation coefficient μ , which we have taken to be constant, is not only somewhat material dependent but is also beam energy dependent. If the beam is truly monochromatic this is not a problem. However, for a polychromatic beam the transmitted total energy will depend on the *distribution* of mass along a path, not just the total mass.
- **Other Attenuation Mechanisms.** We have included only Compton scattering in the model. Four other mechanisms (outlined in the supplementary material) contribute to the attenuation. While Compton scattering is the dominant contributor, photoelectric scattering will have some effect. It becomes important at the lower energies of interest and for materials of relatively high atomic number – such as calcium which is concentrated in bone. The major effect of ignoring photoelectric scattering is quantitative mass uncertainty.
- There are a number of **Detector-Related Effects** which affect radiograph accuracy. Energy detection efficiency can be a function of beam intensity, photon energy and even pixel location. Detectors are subject to point-spread effects in which even an infinitesimally narrow beam results in a finitely narrow detection spot. These types of effects are usually well-understood, documented and can be corrected for in the data. Detection is also prone to noise from secondary radiation, background radiation, or simply manufacturing variances.

8.9 Additional Resources

There is a variety of online source material that expands on any of the material presented here. Here are a few starting points.

https://www.nde-ed.org/EducationResources/CommunityCollege/Radiography/c_rad_index.htm

http://web.stanford.edu/group/glam/xlab/MatSci162_172/LectureNotes/01_Properties%20&%20Safety.pdf

http://radiologymasterclass.co.uk/tutorials/physics/x-ray_physics_production.html

Several physical processes contribute to absorption and scattering of individual photons as they pass through matter. Collectively, these processes alter the geometry and intensity of a beam of such photons. What follows is a brief description of each.

- The **Photoelectric Effect** is the absorption of an x-ray photon by an atom accompanied by the ejection of an outer-shell electron. This ionized atom then re-absorbs an electron and emits an x-ray of energy characteristic of the atom. The daughter x-ray is a low-energy photon which is quickly re-absorbed and is effectively removed from the x-ray beam. Photoelectric absorption is the dominant process for photon energies below about 100keV and when interacting with materials of high atomic number.
- **Rayleigh Scattering** is the process of a photon interacting with an atom without energy loss. The process is similar to the collision of two billiard balls. Rayleigh scattering is never the dominant mechanism, at any energy.
- **Compton Scattering** occurs when an x-ray photon interacts with an electron imparting some energy to the electron. Both electron and photon are emitted and the photon undergoes a directional change or scatter. Compton Scattering is the dominant process for soft tissue at photon energies between about 100keV through about 8MeV.
- **Pair Production** is the process in which a photon is absorbed producing a positron-electron pair. The positron quickly decays into two 510keV x-ray photons. Pair production is only significant for photon energies of several MeV or more.
- **Photodisintegration** can occur for photons of very high energy. Photodisintegration is the process of absorption of the photon by an atomic nucleus and the subsequent ejection of a nuclear particle.

Exercises

1. Given $T_{kj} = 0.42$, what does this value mean?
2. What is accomplished in the matrix multiply when multiplying an object vector by the k th row of T ?
3. Explain how you change the radiographic operator when changing from one view to many.
4. Why do we use pixel values b_k where b_k is defined as

$$b_k \equiv -\alpha \ln \left(\frac{p_k}{p_0} \right)$$

instead of the expected photon count p_k ?

9 Radiography for Linear Algebra Lab #2

9.1 Radiographic Scenarios and Notation

A single-view radiographic setup consists of an area of interest where the object will be placed, and a single screen onto which the radiograph will be recorded. A multiple-view radiographic setup consists of a single area of interest experimentally designed so that radiographs of this area can be recorded for different locations about the object.

The geometry of radiographic scenarios is illustrated in figures 1 and 2. The notation is as follows.

- Slice of region of interest: n by n array of voxels.
- Total number of voxels in each slice is $N = n^2$.
- Each voxel has a width and height of 1 unit.
- For each radiographic view we record m pixels of data.
- The width of each pixel is $ScaleFac$. If $ScaleFac = 1$ then pixel width is the same as voxel width.
- Number of radiographic angles (views): a .
- Total number of pixels in the radiograph: $M = am$
- Angle of the i^{th} view (measured in degrees east of south): θ_i
- Object mass at voxel j is x_j
- Recorded radiograph value at pixel k is b_k

In this Lab we will be constructing matrix representations of the radiographic transformation operators for example scenarios. Recall that an object can be represented by an N -dimensional vector x and a set of radiographs can be represented by an M -dimensional vector b . What will be the size of the corresponding matrix operator that maps object vector x to radiograph vector b ? Read again the definition of the matrix operator from the Radiography lecture notes. Notice that particular values for x and b are *not necessary* for computing the matrix of the radiographic transformation!

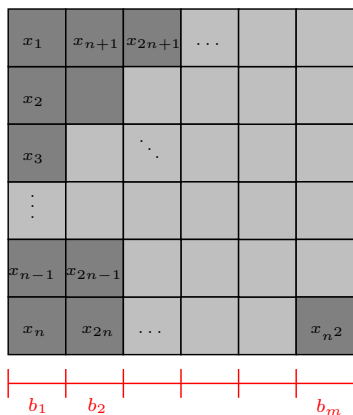


Figure 1: The geometry of a single view radiographic transformation.

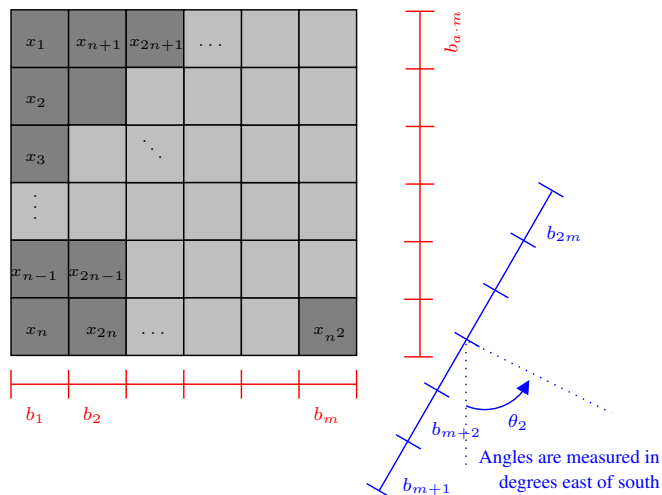
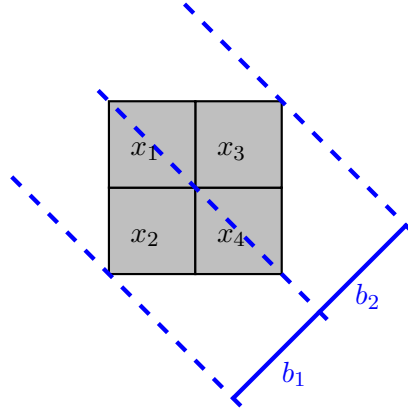


Figure 2: The Geometry of a multiple view radiographic transformation showing view 1, view 2, and view a .

9.2 A First Example

Let's look at a specific scenario. For the setup pictured below, we have:

- Total number of voxels: $N = 4$ ($n = 2$).
- Total number of pixels: $M = m = 2$
- $ScaleFac = \sqrt{2}$
- Number of views: $a = 1$
- Angle of the single view: $\theta_1 = 45^\circ$



Recalling that T_{kj} is the fraction of voxel j which projects perpendicularly onto pixel k , the matrix associated with this radiographic setup is $T = \begin{bmatrix} 1/2 & 1 & 0 & 1/2 \\ 1/2 & 0 & 1 & 1/2 \end{bmatrix}$. Be sure and check this to see if you agree! Hence, for *any* input vector x , the radiographic output is $b = Tx$.

Find the output when the object is the vector $x = \begin{bmatrix} 10 \\ 0 \\ 5 \\ 10 \end{bmatrix}$.

9.3 Now it's your turn!

Note: Some of the following exercises will ask you to use Matlab or Octave to compute radiographic transformations. You will need to first download the function `tomomap.m` from the class website. If you do not yet have access to personal or department computing resources, you can complete this assignment using online Octave in a web browser. Instructions are given at the end of this document.

Instructions for using octave-online.net

1. Download the file `tomomap.m` from the class website to your computer in a location of your choice.
2. Open `octave-online.net` in a web browser. Close (or read and close) the introduction splash window.
3. Click the three bars in the upper right of the screen and sign in.
4. The far left window is a list of your files. You need to place the file `tomomap.m` into this octave-online folder. To do this either (a) drag and drop `tomomap.m` onto this window or (b) click the upload file icon in the window, find `tomomap.m`, and click open. The file opens automatically in the editor window.
5. FYI: The rightmost window is the command prompt terminal. You will type your commands here. You also have the option of creating your own scripts and functions to save and run.

6. Run this example call to function `tomomap.m` by typing the following at the command prompt and hitting return.

```
T = tomomap(2,2,[20 60],1)
```

You should see twelve entries in the variable T . All other entries are zero. To see the full array try the following:

```
T = full(tomomap(2,2,[20,60],1))
```

7. Now you are ready to complete the assignment. You can always read the comments at the beginning of `tomomap.m` (see editor window) to understand more about what this code does and how to use it.

9.3.1 Part I: Finding the matrix associated with a radiographic transformation, and using the matrix to determine radiographic outputs.

1. Suppose you have the setup where

- Height and width of image in voxels: $n = 2$ (Total voxels $N = 4$)
- Pixels per view in radiograph: $m = 2$
- $ScaleFac = 1$
- Number of views: $a = 2$
- Angle of the views: $\theta_1 = 0^\circ, \theta_2 = 90^\circ$

- (a) Sketch this setup.
 (b) Calculate the matrix associated with the setup.
 (c) Find the radiographs of the following objects.

4	3
6	1

0	1
1	0

2. Suppose you have the setup where

- Height and width of image in voxels: $n = 2$ (Total voxels $N = 4$)
- Pixels per view in radiograph: $m = 2$
- $ScaleFac = 1$
- Number of views: $a = 2$
- Angle of the views: $\theta_1 = 0^\circ, \theta_2 = 45^\circ$

- (a) Sketch this setup.
 (b) Calculate the matrix associated with the setup.
 (c) Repeat step (b) using the code `tomomap`.

3. Suppose you have the setup where

- Height and width of image in voxels: $n = 2$ (Total voxels $N = 4$)
- Pixels per view in radiograph: $m = 2$
- $ScaleFac = \sqrt{2}$
- Number of views: $a = 2$
- Angle of the views: $\theta_1 = 45^\circ, \theta_2 = 135^\circ$

(a) Sketch this setup.

(b) Calculate the matrix associated with the setup.

(c) Repeat step (b) using the code `tomomap`.

4. Suppose you have the setup where

- Height and width of image in voxels: $n = 2$ (Total voxels $N = 4$)
- Pixels per view in radiograph: $m = 4$
- $ScaleFac = \sqrt{2}/2$
- Number of views: $a = 1$
- Angle of the views: $\theta_1 = 45^\circ$

(a) Sketch this setup.

(b) Calculate the matrix associated with the setup.

(c) Repeat step (b) using the code `tomomap`.

5. Suppose you have the setup where

- Height and width of image in voxels: $n = 4$ (Total voxels $N = 16$)
- Pixels per view in radiograph: $m = 2$
- $ScaleFac = 1$
- Number of views: $a = 2$
- Angle of the views: $\theta_1 = 0^\circ, \theta_2 = 45^\circ$

(a) Sketch this setup.

(b) Calculate the matrix associated with the setup.

(c) Find the radiographs of images A, B, and C from lab 1 under this transformation.

(d) Repeat steps (b) and (c) using the code `tomomap`.

6. Suppose you have the setup where

- Height and width of image in voxels: $n = 4$ (Total voxels $N = 16$)
- Pixels per view in radiograph: $m = 4$
- $ScaleFac = 1$
- Number of views: $a = 3$
- Angle of the views: $\theta_1 = 0^\circ, \theta_2 = 25.5^\circ, \text{ and } \theta_3 = 90^\circ$

(a) Sketch this setup.

(b) Calculate the matrix associated with the setup using `tomomap`.

(c) Find the radiographs of images A, B, and C from lab 1 under this transformation.

9.3.2 Part II: Radiographs of Linear Combinations of Objects

Take the two objects in exercise 1 above to be x (left object) and y (right object). For each of the transformations in exercises 1, 3, and 4 from Part I answer the following questions.

1. What is the radiograph of the object $3x$?
2. What is the radiograph of the object $0.5y$?
3. What is the radiograph of the object $3x + 0.5y$?

4. Generalize these observations to arbitrary linear combinations of object vectors. Write your conjecture(s) in careful mathematical notation.

10 Transformations

First, let us clear up some terminology. Transformation is just another word for function. We use the word to suggest that we are working in spaces (in this class vector spaces) and the function will transform the space to another space. In a multivariable calculus class, we see this when thinking about about a change of variables. For example, we transform rectangular coordinates to polar through a transformation. In Figure 7 we see that this change of variables transformation takes round things and makes them rectangular.

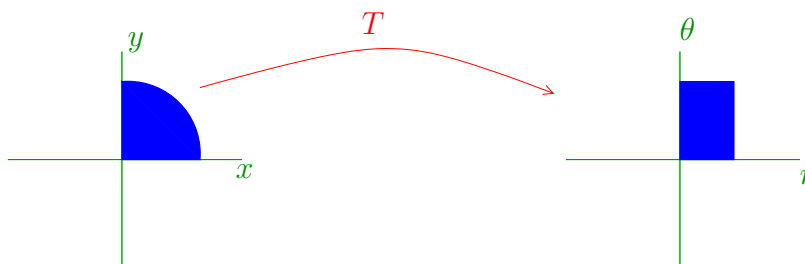


Figure 7:

Another type of transformation is when we project from one space to a lower dimensional space. For example, we can project a points in \mathbb{R}^2 to their component along the y axis as in Figure 8. In this transformation, we take a space that is a 2-d plane and transform it into a 1-d line.

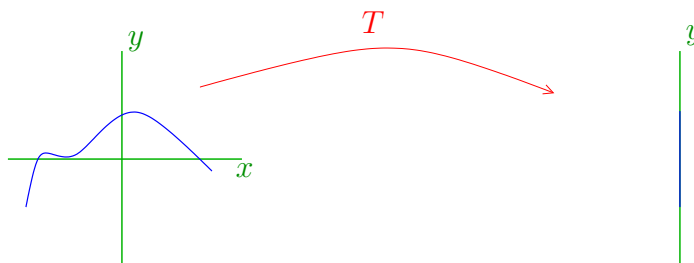


Figure 8:

Notice that in Lab #2 above, we found that when we apply a radiographic transformation to a linear combination, $\alpha u + \beta v$, we get a linear combination of radiographs, $\alpha Tu + \beta Tv$

out. This is property of the radiographic transformation is useful because we may wonder what is an object made of. If we know part of what is being radiographed (and what the corresponding radiograph should be), we can subtract that away to be able to view the radiograph of the part we don't know.

Example 10.1. *For example, suppose we expect an object to look like what we see on the left in Figure 9 but the truth is actually like we see on the right in Figure 9. We then expect*

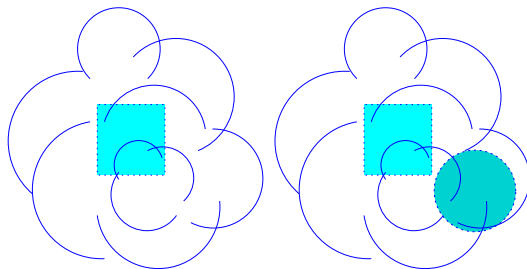


Figure 9:

that the radiograph that comes from an object, x_{exp} like we see on the left in Figure 9 to be a certain radiograph, call it b_{exp} . But, we take a radiograph of the actual object, x , and we get the radiograph b . Now, we know for sure that the expected parts are in the object. We can then remove b_{exp} from the radiograph so that we can dig deeper into what else is present. Thus, we want to know about the radiograph, b_{unexp} , of the unexpected object, call this x_{unexp} , that is present. So we compute this radiograph like this:

$$T(x_{unexp}) = T(x - x_{exp}) = b - b_{exp}.$$

Another reason this might be helpful comes in possible changes in an object.

Example 10.2. *Suppose you radiograph an object as in Figure 9 and find that the radiograph is b_{exp} , but weeks later, you radiograph the same object (or so you think) and you get a radiograph that is 1.3 times the radiograph b_{exp} . This could mean that the object now looks more like one of the objects we see in Figure 10. Again, we can see that the density is*

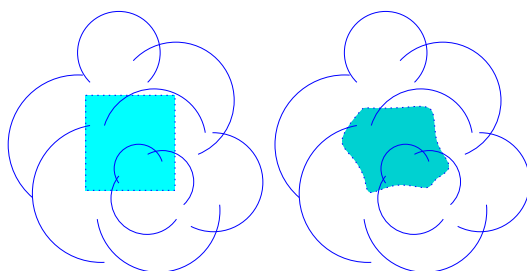


Figure 10:

proportionally larger, possibly meaning the object grew.

10.1 Linear Transformations

Being able to perform the type of investigation above with our radiographic transformation is very useful. It turns out that this property is useful beyond the application of radiography, so useful that we define transformations with this property.

Definition 10.1. Let V and W be vector spaces and let T be a function that maps vectors in V to vectors in W . We say that $T : V \rightarrow W$ is a linear transformation if for vectors $v_1, v_2 \in V$ and scalars α, β ,

$$T(\alpha v_1 + \beta v_2) = \alpha T(v_1) + \beta T(v_2).$$

Note: We also call T a homomorphism. (Linear and homomorphism mean the same thing in this context.)

Notice, also, that the radiographic transformation is an example of a linear transformation. Let's look at some more examples.

Example 10.3. Consider the vector space $V = \mathcal{M}_{2 \times 3}$. Let $T : V \rightarrow W$ be defined as follows.

$$T \begin{pmatrix} a & b & c \\ d & f & g \end{pmatrix} = \begin{pmatrix} a + b + c \\ d + f + g \end{pmatrix}.$$

We can determine whether T is linear by determining whether or not $T(\alpha v_1 + \beta v_2) = \alpha T(v_1) + \beta T(v_2)$ for some $v_1, v_2 \in V$ and scalars α and β . Indeed, if $v_1, v_2 \in V$, then

$$v_1 = \begin{pmatrix} a_1 & b_1 & c_1 \\ d_1 & f_1 & g_1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} a_2 & b_2 & c_2 \\ d_2 & f_2 & g_2 \end{pmatrix}.$$

Thus,

$$\begin{aligned} T(\alpha v_1 + \beta v_2) &= T \begin{pmatrix} \alpha a_1 + \beta a_2 & \alpha b_1 + \beta b_2 & \alpha c_1 + \beta c_2 \\ \alpha d_1 + \beta d_2 & \alpha f_1 + \beta f_2 & \alpha g_1 + \beta g_2 \end{pmatrix} \\ &= \begin{pmatrix} \alpha a_1 + \beta a_2 + \alpha b_1 + \beta b_2 + \alpha c_1 + \beta c_2 \\ \alpha d_1 + \beta d_2 + \alpha f_1 + \beta f_2 + \alpha g_1 + \beta g_2 \end{pmatrix} \\ &= \begin{pmatrix} \alpha(a_1 + b_1 + c_1) + \beta(a_2 + b_2 + c_2) \\ \alpha(d_1 + f_1 + g_1) + \beta(d_2 + f_2 + g_2) \end{pmatrix} \\ &= \begin{pmatrix} \alpha(a_1 + b_1 + c_1) \\ \alpha(d_1 + f_1 + g_1) \end{pmatrix} + \begin{pmatrix} \beta(a_2 + b_2 + c_2) \\ \beta(d_2 + f_2 + g_2) \end{pmatrix} \\ &= \alpha \begin{pmatrix} a_1 + b_1 + c_1 \\ d_1 + f_1 + g_1 \end{pmatrix} + \beta \begin{pmatrix} a_2 + b_2 + c_2 \\ d_2 + f_2 + g_2 \end{pmatrix} \\ &= \alpha T(v_1) + \beta T(v_2). \end{aligned}$$

Thus, T maps linear combinations to linear combinations with the same scalars. So, T is a linear transformation.

Example 10.4. In basic algebra classes, we learn that $f(x) = mx + b$ is a linear function. Let's explore this function in this example. First notice that $f : \mathbb{R} \rightarrow \mathbb{R}$. That is, f maps real numbers to real numbers. We also know that \mathbb{R} is a vector space. So, is f linear like we learned to call it? The answer is no. Here's why. Let $x, y \in \mathbb{R}$ and let α, β be scalars. Then

$$f(\alpha x + \beta y) = m(\alpha x + \beta y) + b = \alpha(mx) + \beta(my) + b \neq \alpha(mx + b) + \beta(my + b).$$

Thus, f does not satisfy our definition of linear. In fact, mathematicians prefer to call functions of the form $f(x) = mx + b$ affine.

So, we should explore what went wrong with the last function. It seems that the constant, b caused problems. Notice that if $b = 0$ above, we would have the function $f(x) = mx$ and

$$f(\alpha x + \beta y) = m(\alpha x + \beta y) + b = \alpha(mx) + \beta(my) = \alpha f(x) + \beta f(y).$$

That means, $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(x) = mx$ is linear. There's a hidden property of linear transformations that we should expose now.

Theorem 10.1. *Let V and W be vector spaces. If $T : V \rightarrow W$ is a linear transformation, then $T(0) = 0$.*

Proof: Let V and W be vector spaces and let $T : V \rightarrow W$ be a linear transformation. Notice that $0 \in V$ and $0 \in W$. (Note also that these two vectors called 0 need not be the same vector.) We also know that, for any scalar α , $T(0) = T(\alpha 0) = \alpha T(0)$. We can use this equation to solve for $T(0)$ and we get that either $\alpha = 1$ or $T(0) = 0$. Since we know that α can be any scalar, $T(0) = 0$ must be true. \square

This last theorem gives us a quick check to see whether a mapping is linear. We can determine whether or not 0 maps to 0. If not, T cannot be linear.

Exercises

1. Determine which of the following is a linear transformation. Prove your answer.

(a) Define $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ by $f(v) = Mv + x$, where

$$M = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix} \quad \text{and} \quad x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

(b) Define $\mathcal{F} : V \rightarrow \mathcal{P}_1$, where

$$V = \{ax^2 + (3a - 2b)x + b \mid a, b \in \mathbb{R}\} \subseteq \mathcal{P}_2.$$

by

$$\mathcal{F}(ax^2 + (3a - 2b)x + b) = 2ax + 3a - 2b.$$

(c) Define $\mathcal{G} : \mathcal{P}_2 \rightarrow \mathcal{M}_{2 \times 2}$ by

$$\mathcal{G}(ax^2 + bx + c) = \begin{pmatrix} a & a - b \\ c - 2 & c + 3a \end{pmatrix}.$$

(d) Define $h : V \rightarrow \mathcal{P}_1$, where

$$V = \left\{ \begin{pmatrix} a & b & c \\ 0 & b - c & 2a \end{pmatrix} \mid a, b, c \in \mathbb{R} \right\} \subseteq \mathcal{M}_{2 \times 3}$$

by

$$h \begin{pmatrix} a & b & c \\ 0 & b - c & 2a \end{pmatrix} = ax + c.$$

(e) Define $f : \mathcal{I} \rightarrow \mathcal{P}_2$ where $\mathcal{I} = \{I \mid I \text{ is of the form below}\}$ by

$$f(I) = ax^2 + (b + c)x + (a + c),$$

where I is the image below

$$I = \begin{array}{|c|c|c|} \hline & 3a & \\ \hline a-b & & a \\ \hline & b+a & \\ \hline 2b-c & & b \\ \hline & c+b & \\ \hline \end{array}$$

(f) Define $f : \mathcal{M}_{2 \times 2} \rightarrow \mathbb{R}^4$ by

$$f \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}.$$

(g) Define $f : \mathcal{P}_2 \rightarrow \mathbb{R}^2$ by

$$f(ax^2 + bx + c) = \begin{pmatrix} a + b \\ a - c \end{pmatrix}.$$

(h) Let \mathcal{H}_{14} be the set of all possible heat states representing the temperature difference, from the end temperatures, every 1 cm along a 6 cm long rod (start measuring at one end). Define a function $h : \mathcal{H}_6 \rightarrow \mathcal{H}_6$ by averaging neighboring heat coordinates in the heat state. That is, if $v = (0, a_1, a_2, a_3, a_4, a_5, 0)$ is a heat state in \mathcal{H}_6 then

$$\begin{aligned} & h(0, a_1, a_2, a_3, a_4, a_5, 0) \\ &= \left(0, \frac{0 + a_1 + a_2}{3}, \frac{a_1 + a_2 + a_3}{3}, \frac{a_2 + a_3 + a_4}{3}, \frac{a_3 + a_4 + a_5}{3}, \frac{a_4 + a_5 + 0}{3}, 0 \right). \end{aligned}$$

Show h is a linear transformation.

11 Radiography and Tomography in Linear Algebra Lab #3

In this activity, you will explore some of the properties of radiographic transformations. In Lab #2 you found six radiographic transformation operators. These operators were defined by each of the six setups given in Lab #2. You then found the corresponding matrix T so that, mathematically, we defined the transformations as a matrix multiply. Then you applied the matrix multiply to specific object images. The object image consisted of four voxels and was represented as a vector in \mathbb{R}^4 . The radiographic image was represented as a vector in \mathbb{R}^M , one entry for each radiographic pixel. The price we pay for this representation is that we no longer have the geometry of the radiographic setup encoded in the representation. The use of representations in \mathbb{R}^n is a computational tool and not geometrically

descriptive of vector spaces of images. We want to reiterate that this is *only* a representation and that these images are **not** vectors in \mathbb{R}^M . Because of these new image representations, each transformation could be constructed as a matrix operator in $\mathcal{M}_{M \times 4}(\mathbb{R})$.

In this lab, we are interested, specifically, in the radiographic transformations 1, 3, and 4 from Lab # 2 and not the specific object image examples given in that lab.

Task 1

For each of the radiographic transformations (defined by the radiographic setups) 1,3 and 4 which you found in Lab #2 answer the following questions. Justify your conclusions.

1. Is it possible for two different objects to produce the same radiograph? If so, give an example.
2. Are any nonzero objects invisible to this operator? If so, give an example. We say that an object is nonzero if not all entries are zero. We say that an object is invisible if it produces the zero radiograph.
3. Are there radiographs (in the appropriate dimension for the problem) that cannot be produced as the radiograph of *any* object? If so, give an example.

Task 2

Go a little deeper into understanding the operators 1, 3, and 4 from Lab #2 by answering these questions about operators 1,3, and 4.

1. Choose any two objects that produce the same radiograph and subtract them. What is special about the resulting object?
2. Describe the set of all invisible objects. This could involve an equation that the entries would have to satisfy or a few specific objects that could be used to construct all other such objects. Be creative.
3. Describe the set of radiographs that can be produced from all possible objects. This may require similar creativity.

Task 3

Now put this more into the language of Linear Algebra. For each of the operators 1, 3, and 4 from Lab # 2 do the following.

1. Give a basis for the set of all invisible objects.
2. Give a basis for the set of all possible radiographs.

12 Properties of Transformations

In Lab # 3, we saw several properties of linear transformations that were useful to recognize. We saw that it was possible for two objects to produce the same radiograph. This can be an issue in the case of brain radiography. We would want to know if it was possible for an abnormal brain to produce the same radiograph as a normal brain. We also saw that it was possible to have radiographs that could not be produce from any object. This becomes important in being able to recognize noise or other corruption in a given radiograph.

12.1 Injections and Surjections

Again, it turns out that these properties are not only important in radiography. There are many other scenarios (some application based and some theoretical) where we need to know if a transformation has these properties. So, let's define them.

Definition 12.1. Let V and W be vector spaces. We say that the transformation $T : V \rightarrow W$ is injective if the following property is true:

Whenever $T(u) = T(v)$ it must be true that $u = v$.

A transformation that is injective is also said to be one-to-one (1-1) or an injection.

Notice, in Figure 11, we see that the transformation represented on the left is 1-1, but the transformation represented on the right is not because both v_1 and v_2 map to w_4 , but $v_1 \neq v_2$.

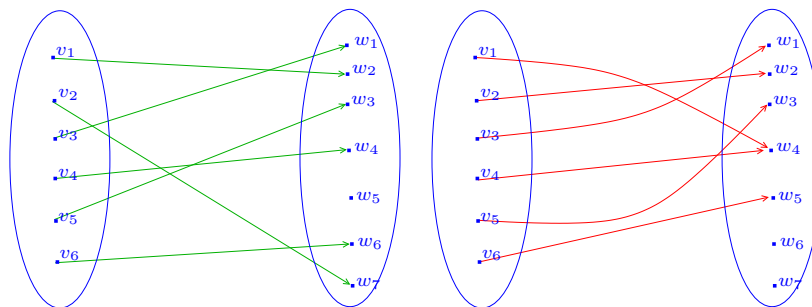


Figure 11:

Let us now look at some more concrete examples.

Example 12.1. Let $T : \mathcal{P}_2 \rightarrow \mathcal{M}_{3 \times 2}$ be defined by

$$T(ax^2 + bx + c) = \begin{pmatrix} a & b & c \\ 0 & 0 & 0 \end{pmatrix}.$$

Notice that T is a linear transformation. Indeed, let $v = ax^2 + bx + c$ and $u = fx^2 + gx + h$. Then

$$\begin{aligned} T(\alpha v + \beta u) &= T((\alpha a + \beta f)x^2 + (\alpha b + \beta g)x + (\alpha c + \beta h)) \\ &= \begin{pmatrix} \alpha a + \beta f & \alpha b + \beta g & \alpha c + \beta h \\ 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \alpha a & \alpha b & \alpha c \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \beta f & \beta g & \beta h \\ 0 & 0 & 0 \end{pmatrix} \\ &= \alpha T(v) + \beta T(u). \end{aligned}$$

Thus, T is a linear transformation.

Definition 12.2. Let V and W be vector spaces. We say that the transformation $T : V \rightarrow W$ is surjective if every element in W is mapped to. That is, if $w \in W$, then there exists a $v \in V$ so that $T(v) = w$.

A transformation that is surjective is also said to be onto or a surjection.

In Figure 11, neither of the transformations represented are surjective. On the left, we see that no vector maps to w_5 . On the right, no vector maps to w_6 nor does a vector map to w_7 . In Figure 12, the left two are surjective and neither is injective. Again, let's consider some

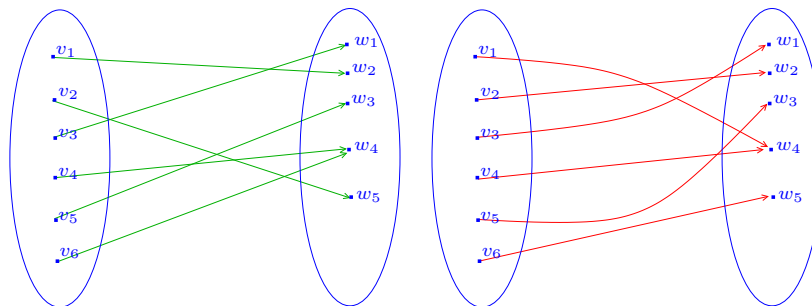


Figure 12:

concrete examples.

Example 12.2. Let $T : \mathbb{R}^2 \rightarrow \mathcal{M}_{3 \times 2}$ be defined by

$$T \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a+b \\ 0 & -a \end{pmatrix}.$$

We can show that T is a linear transformation. (I'll leave this for the reader to do.) We want to know if T is one-to-one and/or onto. Notice that if

$$T \begin{pmatrix} a \\ b \end{pmatrix} = T \begin{pmatrix} c \\ d \end{pmatrix}$$

Then

$$\begin{pmatrix} a & -b \\ b & a+b \\ 0 & -a \end{pmatrix} = \begin{pmatrix} c & -d \\ d & c+d \\ 0 & -c \end{pmatrix}$$

Matching up entries, gives us $a = c$, $-b = -d$, $b = d$, $a + b = c + d$, $0 = 0$, and $-c = -d$. Thus,

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}.$$

So, by definition, T is one-to-one.

Notice that,

$$w = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \in \mathcal{M}_{3 \times 2}.$$

But, there is no $v \in \mathbb{R}^2$ so that $T(v) = w$ because every output has a 0 in the (3, 1) entry.

Example 12.3. Let $T : \mathbb{R}^2 \rightarrow \mathcal{P}_1$ be defined by

$$T \begin{pmatrix} a \\ b \end{pmatrix} = ax + b.$$

Notice that if we pick $w \in \mathcal{P}_1$, then $w = ax + b$ for some $a, b \in \mathbb{R}$. Notice also that if we let

$$v = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$$

then $T(v) = w$. Thus, T is onto.

Now, suppose

$$T \begin{pmatrix} a \\ b \end{pmatrix} = T \begin{pmatrix} c \\ d \end{pmatrix}.$$

Then

$$ax + b = cx + d.$$

Matching up like terms gives us that $a = c$ and $b = d$. That is

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}.$$

So, T is one-to-one.

12.2 Bijections and Isomorphisms

We see above that sometimes a linear transformation can be both injective and surjective. In this subsection we discuss this special type of linear transformation.

Definition 12.3. We say that a linear transformation, $T : V \rightarrow W$, is bijjective if T is both injective and surjective. We call a bijective transformation a bijection or an isomorphism.

Definition 12.4. Let V and W be vector spaces. If there exists a bijection mapping between V and W , then we say that V is isomorphic to W and we write $V \cong W$.

Notice that in the above example we found a bijection. This means that $\mathcal{P}_1 \cong \mathbb{R}^2$. Notice also that $\dim \mathcal{P}_1 = \dim \mathbb{R}^2$. This is not a coincidence. In what follows, we will see the reason this happens. But first, a preliminary result.

Lemma 12.1. Let V and W be vector spaces. Let $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ be a basis for V . $T : V \rightarrow W$ is an injective linear transformation if and only if $\{T(v_1), T(v_2), \dots, T(v_n)\}$ is a linearly independent set in W .

Proof: As with every proof about linear dependence/independence, we will assume the following linear dependence relation is true. Let $\alpha_1, \alpha_2, \dots, \alpha_n$ be scalars so that

$$\alpha_1 T(v_1) + \alpha_2 T(v_2) + \dots + \alpha_n T(v_n) = 0.$$

Then because T is linear, we know that

$$T(\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n) = 0.$$

But, we also know that $T(0) = 0$. That means that

$$T(\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n) = T(0).$$

And, since T is 1-1, we know that (by definition)

$$\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = 0.$$

Finally, since \mathcal{B} is a basis for V , \mathcal{B} is linearly independent. Thus,

$$\alpha_1 = \alpha_2 = \dots = \alpha_n = 0.$$

Thus, $\{T(v_1), T(v_2), \dots, T(v_n)\}$ is linearly independent.

Now suppose that T is linear, let $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ be a basis for V , and suppose $\{T(v_1), T(v_2), \dots, T(v_n)\} \subset W$ is linearly independent. Suppose $u, v \in V$ so that $T(u) = T(v)$. So, $T(u - v) = 0$. Since $u, v \in V$, there are scalars $\alpha_1, \alpha_2, \dots, \alpha_n$ and $\beta_1, \beta_2, \dots, \beta_n$ so that

$$u = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n \quad \text{and} \quad v = \beta_1 v_1 + \beta_2 v_2 + \dots + \beta_n v_n.$$

Thus

$$T((\alpha_1 - \beta_1)v_1 + (\alpha_2 - \beta_2)v_2 + \dots + (\alpha_n - \beta_n)v_n) = 0.$$

This leads us to the linear dependence relation

$$(\alpha_1 - \beta_1)T(v_1) + (\alpha_2 - \beta_2)T(v_2) + \dots + (\alpha_n - \beta_n)T(v_n) = 0.$$

Since $\{T(v_1), T(v_2), \dots, T(v_n)\}$ is linearly independent, we know that

$$\alpha_1 - \beta_1 = \alpha_2 - \beta_2 = \dots = \alpha_n - \beta_n = 0.$$

That is, $u = v$. So, T is injective. \square .

Notice that Lemma 12.1 tells us that if V is n -dimensional then basis elements of V map to basis elements of an n -dimensional subspace of W . In particular, if $\dim W = n$ also, then we see a basis of V maps to a basis of W . This is very useful in proving the next result.

Theorem 12.1. *Let V and W be (finite dimensional) vector spaces. $V \cong W$ if and only if $\dim V = \dim W$.*

Note: If we allow the dimension of a vector space to be infinite, then this is not necessarily true. In this class, we restrict to discussions of finite dimensional vector spaces only, so we are happy with this result.

Proof: Suppose $\dim V = \dim W$. Suppose also that a basis for V is $\mathcal{B}_V = \{v_1, v_2, \dots, v_n\}$ and a basis for W is $\mathcal{B}_W = \{w_1, w_2, \dots, w_n\}$. Then we can define $T : V \rightarrow W$ to be the linear transformation so that

$$T(v_1) = w_1, T(v_2) = w_2, \dots, T(v_n) = w_n.$$

We will show that T is an isomorphism. Now, we know that if $w \in W$, then $w = \alpha_1 w_1 + \alpha_2 w_2 + \dots + \alpha_n w_n$ for some scalars $\alpha_1, \alpha_2, \dots, \alpha_n$. We also know that $v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n \in V$. Since T is linear, we can see that

$$\begin{aligned} T(v) &= T(\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n) \\ &= \alpha_1 T(v_1) + \alpha_2 T(v_2) + \dots + \alpha_n T(v_n) \\ &= \alpha_1 w_1 + \alpha_2 w_2 + \dots + \alpha_n w_n = w. \end{aligned}$$

Thus, T is onto. Now, suppose that $T(v) = T(u)$ where $v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$ and $u = \beta_1 v_1 + \beta_2 v_2 + \dots + \beta_n v_n$ are vectors in V . Then we have

$$\begin{aligned} T(\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n) &= T(\beta_1 v_1 + \beta_2 v_2 + \dots + \beta_n v_n) \\ \alpha_1 T(v_1) + \alpha_2 T(v_2) + \dots + \alpha_n T(v_n) &= \beta_1 T(v_1) + \beta_2 T(v_2) + \dots + \beta_n T(v_n) \\ \alpha_1 w_1 + \alpha_2 w_2 + \dots + \alpha_n w_n &= \beta_1 w_1 + \beta_2 w_2 + \dots + \beta_n w_n \\ (\alpha_1 - \beta_1)w_1 + (\alpha_2 - \beta_2)w_2 + \dots + (\alpha_n - \beta_n)w_n &= 0. \end{aligned}$$

Notice that this last equation is a linear dependence relation for the basis \mathcal{B}_W . Since \mathcal{B}_W is linearly independent, we know that

$$\begin{aligned} \alpha_1 - \beta_1 &= 0 \\ \alpha_2 - \beta_2 &= 0 \\ &\vdots \\ \alpha_n - \beta_n &= 0. \end{aligned}$$

That is to say $u = v$. Thus, T is injective. And, therefore, since T is both injective and surjective, T is an isomorphism. Now, since there is an isomorphism between V and W , we know that $V \cong W$.

Now we will prove the other direction. That is, we will show that if $V \cong W$ then $\dim V = \dim W$. First, let us assume that $V \cong W$. This means that there is an isomorphism, $T : V \rightarrow W$, mapping between V and W .

Suppose, for the sake of contradiction, that $\dim V \neq \dim W$. Without loss of generality, assume $\dim V > \dim W$. (We can make this assumption because we can just switch V 's and W 's in the following argument and argue for the case when $\dim V < \dim W$.) Let $\mathcal{B}_V = \{v_1, v_2, \dots, v_n\}$ be a basis for V and $\mathcal{B}_W = \{w_1, w_2, \dots, w_m\}$ be a basis for W . Then $m < n$. We will show that this cannot be true. Lemma 12.1 tells us that since T is one-to-one, the basis \mathcal{B}_V maps to a linearly independent set $\{T(v_1), T(v_2), \dots, T(v_n)\}$ with n elements. But by Theorem 6.1, we know that this is not possible. Thus, our assumption that $n > m$ cannot be true. Again, the argument that tells us that $n > m$ also cannot be true is very similar with V 's and W 's switched. Thus, $n = m$. That is, $\dim V = \dim W$. \square

Theorem 12.1 gives us a tool for creating isomorphisms, when isomorphisms exist. It also tells us that isomorphisms exist between two vector spaces so long as they have the same dimension.

Example 12.4. Let $V = \mathcal{M}_{2 \times 3}$ and $W = \mathcal{P}_5$. We know that $V \cong W$ because both are 6-dimensional vector spaces. Indeed, a basis for V is

$$\mathcal{B}_V = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}$$

and a basis for W is

$$\mathcal{B}_W = \{1, x, x^2, x^3, x^4, x^5\}.$$

Thus, we can create a bijection T that maps V to W . Using Theorem 12.1 we define T as

follows

$$\begin{aligned}T\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} &= 1 \\T\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} &= x \\T\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} &= x^2 \\T\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} &= x^3 \\T\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} &= x^4 \\T\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} &= x^5.\end{aligned}$$

Notice that if we have any vector $v \in V$, we can find where T maps it to in W . Since $v \in V$, we know there are scalars, a, b, c, d, e, f so that

$$v = a\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} + e\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} + f\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

That is,

$$v = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}.$$

Thus, since T is linear,

$$T(v) = a(1) + b(x) + c(x^2) + d(x^3) + e(x^4) + f(x^5) = a + bx + cx^2 + dx^3 + ex^4 + fx^5.$$

Exercises

For the following exercises use only methods discussed in sections 12 or before. That is, do not read ahead and use methods from later sections.

1. Determine which of the transformations in exercise 1 of Section 10 are injective. Prove your answer.
2. Determine which of the transformations in exercise 1 of Section 10 are surjective. Prove your answer.
3. Create a transformation that maps from \mathbb{R}^3 to \mathcal{P}_1 that is surjective. Prove that your transformation is surjective.
4. Can you create an injective transformation that maps \mathbb{R}^2 to $\mathcal{M}_{2 \times 3}$? If so, create and prove that it is injective.
5. Can you create an surjective transformation that maps \mathbb{R}^2 to $\mathcal{M}_{2 \times 3}$? If so, create and prove that it is surjective.
6. Can you create an injective transformation that maps $\mathcal{M}_{2 \times 3}$ to \mathbb{R}^2 ? If so, create and prove that it is injective.

7. Can you create an surjective transformation that maps $\mathcal{M}_{2 \times 3}$ to \mathbb{R}^2 ? If so, create and prove that it is surjective.
8. Create a bijection between \mathcal{P}_1 and $V = \{(x, y, z) | x + y + z = 0\}$. Prove your transformation is linear and is indeed a bijection. If such a bijection doesn't exist, explain why not.

13 Transformation Spaces

In Lab # 3, we saw more properties of linear transformations that were useful to recognize and that were not discussed in Section 12. Some radiographic transformations cannot “see” certain objects or we say that those objects are “invisible” to the transformation. In this application, we really want to know if something might be present yet invisible to the transformation we are using. In the case of brain scans, it would be most unhelpful if we cannot see certain abnormalities because they are invisible to the radiographic setup. If something we want to look for in a brain scan is invisible to our current setup, we can adjust the setup to “see” the object we are looking for. Say, for example, we know that the object on the right in Figure 13 is invisible to our radiographic transformation. Then when it is present along with what we expect (Figure 9 in Section 10), we get the same radiographic transformation b_{exp} and we might go along our merry way, not knowing that something unexpected is present in our object so that instead of what is seen in Figure 9, we actually have what is on the left in Figure 13.

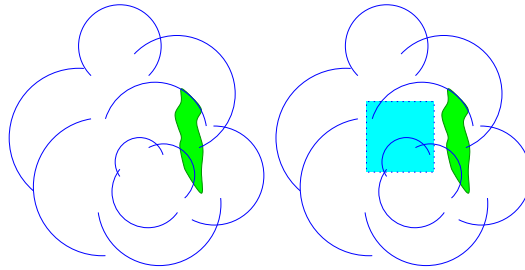


Figure 13:

13.1 The Nullspace

Again, wanting to know which objects are “invisible” to a transformation extends beyond the application of Radiography and Tomography. So, we define the space of all “invisible” objects below.

Definition 13.1. *The nullspace of a linear transformation, $T : V \rightarrow W$, is the subset of V that map to $0 \in W$. That is,*

$$\text{null}(T) = \{v \in V | T(v) = 0\}.$$

We say that the nullity of a linear transformation, T , is the dimension of the subspace $\text{null}(T)$.

Example 13.1. Define $\mathcal{F} : V \rightarrow \mathcal{P}_1$, where

$$V = \{ax^2 + (3a - 2b)x + b \mid a, b \in \mathbb{R}\} \subseteq \mathcal{P}_2.$$

by

$$\mathcal{F}(ax^2 + (3a - 2b)x + b) = 2ax + 3a - 2b.$$

We now find the nullspace of \mathcal{F} :

$$\begin{aligned} \text{null}(\mathcal{F}) &= \{v \in V \mid \mathcal{F}(v) = 0\} \\ &= \{ax^2 + (3a - 2b)x + b \mid 2ax + 3a - 2b = 0\} \\ &= \{ax^2 + (3a - 2b)x + b \mid a = 0, b = 0\} \\ &= \{0\}. \end{aligned}$$

In this case, the nullity is 0 because there are no elements in the basis of $\text{null}\mathcal{F}$

In cases as in Example 13.1, we say that \mathcal{F} has a trivial nullspace, or that the nullspace of \mathcal{F} is trivial.

Example 13.2. Define $h : V \rightarrow \mathcal{P}_1$, where

$$V = \left\{ \begin{pmatrix} a & b & c \\ 0 & b - c & 2a \end{pmatrix} \mid a, b, c \in \mathbb{R} \right\} \subseteq \mathcal{M}_{2 \times 3}$$

by

$$h \left(\begin{pmatrix} a & b & c \\ 0 & b - c & 2a \end{pmatrix} \right) = ax + c.$$

Let's find $\text{null}(h)$.

$$\begin{aligned} \text{null}(h) &= \{v \in V \mid h(v) = 0\} = \left\{ \begin{pmatrix} a & b & c \\ 0 & b - c & 2a \end{pmatrix} \mid h \left(\begin{pmatrix} a & b & c \\ 0 & b - c & 2a \end{pmatrix} \right) = 0, a, b, c \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} a & b & c \\ 0 & b - c & 2a \end{pmatrix} \mid ax + c = 0, a, b, c \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} a & b & c \\ 0 & b - c & 2a \end{pmatrix} \mid a = 0, c = 0, b \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} 0 & b & 0 \\ 0 & b & 0 \end{pmatrix} \mid b \in \mathbb{R} \right\} \\ &= \text{span} \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right\}. \end{aligned}$$

In this case the nullity is 1 because there is one element in the basis for the nullspace.

Note that the above examples are indeed examples of linear transformations (this was proved in Exercise 1 in Section 10).

Note that the name “nullspace” seems to imply that this set is a vector space. In fact, I’ve discussed basis and treated it as if it is a vector space in my examples above. The next theorem justifies my treatment of this set.

Theorem 13.1. Given vector spaces V and W and a linear transformation $T : V \rightarrow W$, the nullspace $\text{null}(T)$ is a subspace of V .

Proof: By definition of $\text{null}(T)$, we know that $\text{null}(T) \subseteq V$. We also know that the zero vector always maps to 0. Thus $0 \in \text{null}(T)$. Now, let α and β be scalars and let $u, v \in \text{null}(T)$. Then $T(u) = 0$ and $T(v) = 0$. Thus, since T is linear, we have

$$T(\alpha u + \beta v) = \alpha T(u) + \beta T(v) = \alpha \cdot 0 + \beta \cdot 0 = 0.$$

So, $\alpha u + \beta v \in \text{null}(T)$. Therefore, $\text{null}(T)$ is a subspace of V . \square

13.2 Domain and Range Spaces

When considering a transformation, we want to know to which vectors we are allowed to apply the transformation. In the case of a Radiographic transformation, we wonder what is the shape and size of objects/images that the particular radiographic transformation uses. This was all part of our radiographic setup. As with most functions, this set is called the domain. In linear algebra, we consider only sets that are vector spaces. So, it is often referred to as the domain space.

There is also an ambient space to which all of the vectors in the domain space map. This ambient space is defined next.

Definition 13.2. We say that the codomain of a linear transformation, $T : V \rightarrow W$, is the ambient vector space to which we map. That is, the codomain of $T : V \rightarrow W$ is W .

In Examples 13.1 and 13.2 the codomain is \mathcal{P}_1 . The codomain tends to be much less interesting for a given problem than the set of all thing mapped to. Notice that not all the vectors in the codomain are mapped to. If they were, then we would say that T is onto.

Definition 13.3. We say that the range space of a linear transformation, $T : V \rightarrow W$, is the subspace of the codomain W that contains all of the outputs from V under the transformation T . That is,

$$\text{ran}(T) = \{T(v) \mid v \in V\}.$$

We say that the rank of a linear transformation, T , is the dimension of the subspace $\text{ran}(T)$.

Notice that, again, in this definition, we imply that the range is a vector space. We will reserve this proof for the Exercises. But, we will state the theorem here.

Theorem 13.2. Let V and W be vector spaces and let $T : V \rightarrow W$ be a linear transformation. Then $\text{ran}(T)$ is a subspace of W .

Proof: Exercise 5

It turns out that in both Examples 13.1 and 13.2 the range is also \mathcal{P}_1 . So in both of these examples, the rank is 2. Like we noticed above, the codomain need not be the range. Let's consider some examples.

Example 13.3. Define $f : \mathcal{M}_{2 \times 2} \rightarrow \mathbb{R}^4$ by

$$f \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a \\ b + a \\ b \\ c \end{pmatrix}.$$

We can find $\text{ran}(f)$ and $\text{null}(f)$. First the range.

$$\begin{aligned} \text{ran}(f) &= \{f(v) \mid v \in \mathcal{M}_{2 \times 2}\} = \left\{ f \left(\begin{array}{cc} a & b \\ c & d \end{array} \right) \mid a, b, c, d \in \mathbb{R} \right\} \\ &= \left\{ \left(\begin{array}{c} a \\ b+a \\ b \\ c \end{array} \right) \mid a, b, c \in \mathbb{R} \right\} \\ &= \text{span} \left\{ \left(\begin{array}{c} 1 \\ 1 \\ 0 \\ 0 \end{array} \right), \left(\begin{array}{c} 0 \\ 1 \\ 1 \\ 0 \end{array} \right), \left(\begin{array}{c} 0 \\ 0 \\ 0 \\ 1 \end{array} \right) \right\}. \end{aligned}$$

Thus $\text{rank}(f) = 3$. Now for the nullspace.

$$\begin{aligned} \text{null}(f) &= \{v \in \mathcal{M}_{2 \times 2} \mid f(v) = 0\} = \left\{ \left(\begin{array}{cc} a & b \\ c & d \end{array} \right) \mid a, b, c, d \in \mathbb{R} \text{ and } f \left(\begin{array}{cc} a & b \\ c & d \end{array} \right) = 0 \right\} \\ &= \left\{ \left(\begin{array}{cc} a & b \\ c & d \end{array} \right) \mid a, b, c, d \in \mathbb{R} \text{ and } \left(\begin{array}{c} a \\ b+a \\ b \\ c \end{array} \right) = 0 \right\} \\ &= \left\{ \left(\begin{array}{cc} a & b \\ c & d \end{array} \right) \mid a, b, c = 0, d \in \mathbb{R} \right\} \\ &= \left\{ \left(\begin{array}{cc} 0 & 0 \\ 0 & d \end{array} \right) \mid d \in \mathbb{R} \right\} \\ &= \text{span} \left\{ \left(\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array} \right) \right\}. \end{aligned}$$

Thus $\text{nullity}(f) = 1$.

Notice that, in this example, the codomain is \mathbb{R}^4 and $\text{ran}(f) \neq \mathbb{R}^4$. That means there are elements in \mathbb{R}^4 that are not mapped to through f . That is, f is not surjective.

Also notice that there is more than one element of the nullspace. That means, since

$$f \left(\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array} \right) = f \left(\begin{array}{cc} 0 & 0 \\ 0 & 2 \end{array} \right),$$

But,

$$\left(\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array} \right) \neq \left(\begin{array}{cc} 0 & 0 \\ 0 & 2 \end{array} \right).$$

Thus, f is not injective.

13.3 Injectivity and Surjectivity Revisited

Let's consider this discussion again from the point of view of radiography. We saw that some transformations had the property where two objects could give the same radiograph. This particular radiographic transformation would not be injective. Notice also, that we found

that if two objects produce the same radiograph, that their difference would then be invisible, that is the difference is in the nullspace of the radiographic transformation. Notice that if there is an object that is invisible to the radiographic transformation, any scalar multiple of it will also be invisible. This means that two different objects are invisible, producing the same radiograph, 0. Thus, the radiographic transformation would not be injective.

Recall, also, that we found that there were radiographs that could not be produced from an object given a certain radiographic transformation. This means that there is a radiograph in the codomain that is not mapped to from the domain. If this happens, the radiographic transformation is not surjective.

We now state the theorems that go along with these results. Our first theorem says that an equivalent statement to

T is onto.

is

The codomain of T is $\text{ran}(T)$.

Theorem 13.3. *Let V and W be vector spaces and let $T : V \rightarrow W$ be a linear transformation. T is surjective if and only if $\text{ran}(T) = W$.*

Proof: Suppose $\text{ran}(T) = W$ then, by definition of $\text{ran}(T)$ if $w \in W$, there is a $v \in V$ so that $f(v) = w$. Thus T is onto. Now, if T is onto, then for all $w \in W$ there is a $v \in V$ so that $T(v) = w$. That means that $W \subseteq \text{ran}(T)$. But, by definition of T and $\text{ran}(T)$, we already know that $\text{ran}(T) \subseteq W$. Thus, $\text{ran}(T) = W$. \square

The next theorem says that an equivalent statement to

T is one-to-one.

is

$\text{null}(T) = \{0\}$.

Theorem 13.4. *A linear transformation, $T : V \rightarrow W$, is injective if and only if $\text{null}(T) = \{0\}$. That is, $T : V \rightarrow \text{ran}(T)$ is a bijection if and only if $\text{null}(T) = \{0\}$.*

Proof: Suppose T is one-to-one and suppose that $u \in \text{null}T$. Then $T(u) = 0$. But, $T(0) = 0$. So, since T is 1-1, we know that $u = 0$. Thus, $\text{null}(T) = \{0\}$. Now, suppose $\text{null}(T) = \{0\}$. We want to show that T is 1-1. Notice that if $u, v \in V$ satisfy

$$T(u) = T(v)$$

then

$$T(u) - T(v) = 0.$$

But since T is linear this gives us that

$$T(u - v) = 0.$$

Thus, $u - v \in \text{null}(T)$. But $\text{null}(T) = \{0\}$. Thus, $u - v = 0$. That is, $u = v$. So, T is 1-1. \square
Theorems 13.3 and 13.4 give us tools to check injectivity and surjectivity. Let's consider a couple examples.

Example 13.4. Define $\mathcal{F} : V \rightarrow \mathcal{P}_1$, where

$$V = \{ax^2 + (3a - 2b)x + b \mid a, b \in \mathbb{R}\} \subseteq \mathcal{P}_2.$$

by

$$\mathcal{F}(ax^2 + (3a - 2b)x + b) = 2ax + 3a - 2b.$$

We showed in Example 13.1 that $\text{null}\mathcal{F} = \{0\}$. Thus \mathcal{F} is injective. We also noted (check that you know how to show this) that $\text{ran}(\mathcal{F}) = \mathcal{P}_1$. Thus, the range and codomain of \mathcal{F} are the same. And, so we know \mathcal{F} is surjective. But now we know that \mathcal{F} is a bijection. This means that $V \cong \mathcal{P}_1$!

Notice also that $\dim V = 2$, $\text{nullity}(\mathcal{F}) = 0$, and $\text{rank}(\mathcal{F}) = 2$.

Example 13.5. Define $h : V \rightarrow \mathcal{P}_1$, where

$$V = \left\{ \begin{pmatrix} a & b & c \\ 0 & b - c & 2a \end{pmatrix} \mid a, b, c \in \mathbb{R} \right\} \subseteq \mathcal{M}_{2 \times 3}$$

by

$$h \left(\begin{pmatrix} a & b & c \\ 0 & b - c & 2a \end{pmatrix} \right) = ax + c.$$

We found that

$$\text{null}h = \text{span} \left\{ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right\}.$$

Thus, h is not 1-1. But, we also noted (again, be sure you know how to show this) that $\text{ran}(h) = \mathcal{P}_1$. Thus, h is onto.

Notice also that $\dim V = 3$, $\text{nullity}(h) = 1$, and $\text{rank}(h) = 2$.

Example 13.6. Define $g : V \rightarrow \mathbb{R}^3$, where $V = \mathcal{P}_1$ by

$$g(ax + b) = \begin{pmatrix} a \\ b \\ a + b \end{pmatrix}.$$

Notice that

$$\begin{aligned} \text{null}(g) &= \{ax + b \mid a, b \in \mathbb{R}, g(ax + b) = 0\} \\ &= \left\{ ax + b \mid a, b \in \mathbb{R}, \begin{pmatrix} a \\ b \\ a + b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\} \\ &= \{ax + b \mid a = 0, b = 0\} \\ &= \{0\} \end{aligned}$$

Thus, g is injective. Now we find the range space.

$$\begin{aligned} \text{ran}(g) &= \{g(ax + b) \mid a, b \in \mathbb{R}\} \\ &= \left\{ \begin{pmatrix} a \\ b \\ a + b \end{pmatrix} \mid a, b \in \mathbb{R} \right\} \\ &= \text{span} \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\} \end{aligned}$$

Notice that since $\text{rank}(g) = 2$ and $\dim \mathbb{R}^3 = 3$, $\mathbb{R}^3 \neq \text{ran}(g)$ and thus g is not onto. Notice also that $\dim V = 2$, $\text{nullity}(g) = 0$, and $\text{rank}(g) = 2$.

13.4 The Rank-Nullity Theorem

In each of the last examples of the previous section, we saw that the following theorem holds:

Theorem 13.5. [*Rank Nullity Theorem*] Let V and W be a vector spaces and let $T : V \rightarrow W$ be a linear transformation. Then the following is true

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

Proof: Let $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$.

First, we consider the case when $\text{ran}(T) = \{0\}$. Then, $\text{ran}(T)$ has no basis, so $\text{rank}(T) = 0$. We also know that if $v \in V$ then $T(v) = 0$. Thus, \mathcal{B} is a basis for $\text{null}(T)$ and $\text{nullity}(T) = n$. Thus, $\text{rank}(T) + \text{nullity}(T) = n$.

Next, we consider the case when $\text{null}(T) = \{0\}$. In this case, $\text{null}(T)$ has no basis so $\text{nullity}(T) = 0$. Now, we refer to Lemma 12.1 and Theorem 13.4. We then know that $\{T(v_1), T(v_2), \dots, T(v_n)\}$ is linearly independent and we also know that $\text{span} \{T(v_1), T(v_2), \dots, T(v_n)\} = \text{ran}(T)$. Thus, $\{T(v_1), T(v_2), \dots, T(v_n)\}$ is a basis for $\text{ran}(T)$ and $\text{rank}(T) = n$. Thus, $\text{rank}(T) + \text{nullity}(T) = n$.

Finally, we consider the case where $\text{rank}(T) = m$ and $\text{nullity}(T) = k$. Let

$$\mathcal{B}_N = \{v_1, v_2, \dots, v_k\}$$

be a basis for $\text{null}(T)$. And let

$$\mathcal{B}_R = \{T(v_{k+1}), T(v_{k+2}), \dots, T(v_{k+m})\}$$

be a basis for $\text{ran}(T)$. Notice that we know that none of the elements of \mathcal{B} are zero (for otherwise this set would not be linearly independent). So, we know that none of $T(v_1), T(v_2), \dots, T(v_k) \in \mathcal{B}_R$. We also know that $\text{span} \{T(v_1), T(v_2), \dots, T(v_n)\} = \text{ran}(T)$ so there must be m linearly independent vectors in $\{T(v_1), T(v_2), \dots, T(v_n)\}$ that form a basis for $\text{ran}(T)$. So, our choice of vectors for the basis of \mathcal{B}_R makes sense.

Our goal is to show that $m+k = n$. That is, we need to show that $\mathcal{B} = \{v_1, v_2, \dots, v_{k+m}\}$. We know that $\{v_1, v_2, \dots, v_{k+m}\} \subseteq \mathcal{B}$. We need to show that if $v \in \mathcal{B}$ then $v \in \{v_1, v_2, \dots, v_{k+m}\}$. Suppose $v \in \mathcal{B}$. Then $T(v) \in \text{ran}(T)$. Suppose $v \notin \text{null}(T)$ (for otherwise, $v \in \mathcal{B}_N$.) Then there are scalars $\alpha_1, \alpha_2, \dots, \alpha_m$ so that

$$T(v) = \alpha_1 T(v_{k+1}) + \alpha_2 T(v_{k+2}) + \dots + \alpha_m T(v_{k+m}).$$

So

$$\alpha_1 T(v_{k+1}) + \alpha_2 T(v_{k+2}) + \dots + \alpha_m T(v_{k+m}) - T(v) = 0.$$

Using that T is linear, we get

$$T(\alpha_1 v_{k+1} + \alpha_2 v_{k+2} + \dots + \alpha_m v_{k+m} - v) = 0.$$

Thus

$$\alpha_1 v_{k+1} + \alpha_2 v_{k+2} + \dots + \alpha_m v_{k+m} - v \in \text{null}(T).$$

So either

$$\alpha_1 v_{k+1} + \alpha_2 v_{k+2} + \dots + \alpha_m v_{k+m} - v = 0$$

or

$$\alpha_1 v_{k+1} + \alpha_2 v_{k+2} + \dots + \alpha_m v_{k+m} - v \in \text{span } \mathcal{B}_N.$$

If

$$\alpha_1 v_{k+1} + \alpha_2 v_{k+2} + \dots + \alpha_m v_{k+m} - v = 0$$

Then $v \in \text{span } \mathcal{B}_T$, but this is only true if $v \in \mathcal{B}_T$ because \mathcal{B} is linearly independent and $v \in \mathcal{B}$ and all the elements of $\{v_{k+1}, v_{k+2}, \dots, v_{k+m}\}$ are also in \mathcal{B} . Now, if

$$\alpha_1 v_{k+1} + \alpha_2 v_{k+2} + \dots + \alpha_m v_{k+m} - v \in \text{span } \mathcal{B}_N$$

then $v \in \mathcal{B}_N$ again because $v \in \mathcal{B}$ and so are all the elements of \mathcal{B}_N .

Thus, $v \in \{v_{k+1}, v_{k+2}, \dots, v_{k+m}\}$. So, we have that $\mathcal{B} = \{v_{k+1}, v_{k+2}, \dots, v_{k+m}\}$. Thus $k+m = n$. That is, $\text{nullity}(T) + \text{rank}(T) = n$. \square

A quick result of this theorem says that we can separate the basis of V into those that map to 0 and those to a basis of $\text{ran}(T)$. More specifically, we have the following result.

Corollary 13.1. *Let V and W be vector spaces and let $T : V \rightarrow W$ be a linear transformation. If $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ is a basis for V , then $\{T(v_1), T(v_2), \dots, T(v_n)\} = \{0\} \cup \mathcal{B}_r$ where \mathcal{B}_r is a basis of $\text{ran}(T)$.*

Proof: We can see in the proof of Theorem 13.5 that \mathcal{B} was split into two sets $\{v_1, v_2, \dots, v_k\}$ (a basis for $\text{null}(T)$) and $\{v_{k+1}, v_{k+2}, \dots, v_n\}$, (where $\{T(v_{k+1}), T(v_{k+2}), \dots, T(v_n)\}$ is a basis for $\text{ran}(T)$). \square

Theorem 13.5 is useful in determining rank and nullity, along with proving results about subspaces. Let's look at an example.

Example 13.7. *Given a linear transformation $T : \mathcal{M}_{2 \times 5} \rightarrow \mathcal{P}_4$. We know that T cannot be one-to-one. The Rank-Nullity Theorem says that $\dim \mathcal{M}_{2 \times 5} = \text{rank}(T) + \text{nullity}(T)$. Since $\text{rank}(T) \leq \dim \mathcal{P}_4 = 5$ and $\dim \mathcal{M}_{2 \times 5} = 10$, we know that $\text{nullity}(T) > 0$. That is, $\text{null}(T) \neq \{0\}$. So by Theorem 13.4, we know that T cannot be injective.*

Notice, we didn't know anything about T except for the spaces from and to which it maps.

Exercises

1. Find the nullspace and nullity of each of the transformations from Exercise 1 in Section 10. Be sure to clearly label which is the nullspace and which is the nullity.
2. For each of the linear transformations in exercise 1 in Section 10, state the domain, V , codomain W , $\dim(W)$, $\dim(V)$, $\text{rank}(T)$ and verify the Rank-nullity Theorem using your answer from exercise 1.
3. For each of the domain spaces, V , specified in exercise 2, find an isomorphism $f : V \rightarrow \mathbb{R}^d$, where $d = \dim(V)$.
4. For each of the codomain spaces, W , specified in exercise 2, find an isomorphism $g : \mathbb{R}^w \rightarrow W$ where $w = \dim(W)$.
5. Prove Theorem 13.2.

14 Matrix Representations

Earlier, we discussed matrices. Matrices tend to be a good way to store information in a computer. They are also, at times, easier to work with (as they were when we solved systems of equations). Coding a linear transformation based on the formula can at times be very tedious also. So, we want to be able to use matrices as a tool for linear transformations as well. Let's look at an example that suggests that this might be a good idea.

Example 14.1. Define $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ by $Tx = Mx$, where M is a 3×2 matrix. T is a linear transformation. We can show this using properties of matrix multiply. Let $x, y \in \mathbb{R}^2$ and let α, β be scalars. Then

$$\begin{aligned} T(\alpha x + \beta y) &= M(\alpha x + \beta y) \\ &= M\alpha x + M\beta y \\ &= \alpha Mx + \beta My \\ &= \alpha Tx + \beta Ty. \end{aligned}$$

This example shows us that matrix multiply defines a linear transformation. So, what if we can define all linear transformations with a matrix multiply? That would be really great! The issue is that we cannot just multiply a vector, in say \mathcal{P}_2 , by a matrix. What would that mean? It turns out that there is a way to turn linear transformations into matrix multiplies! Ok, so I know I have some explaining to do with this one. So, here goes.

What we do know is that, given a basis for V , we are able to represent any vector $v \in V$ as a coordinate vector in \mathbb{R}^n , where $n = \dim V$. Suppose $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ is a basis for V , then we find the coordinate vector $[v]_{\mathcal{B}}$ by finding the scalars, α_i , that make the linear combination $v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$ and we get

$$[v]_{\mathcal{B}} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{R}^n.$$

So, what we want is to figure out how to transform one vector space to another, possibly taking a new path through a matrix multiply. Let's suppose that we call T_1 the transformation that turns vectors in V into vectors in \mathbb{R}^n and call T_2 the transformation that turns vectors in \mathbb{R}^m into vectors in W , where $m = \dim W$. We know that we can multiply vectors in \mathbb{R}^n by $m \times n$ matrices to get vectors in \mathbb{R}^m . So we want to find M that does that and somehow transforms our vectors the way a linear transformation T does (See Figure 14). Recall that Corollary 13.1 tells us that linear transformations take basis elements to basis

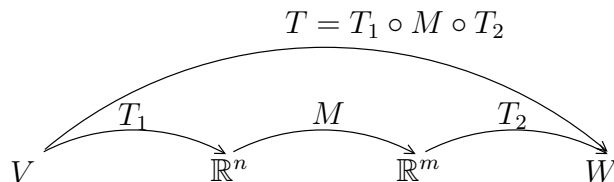


Figure 14:

elements (or to the zero vector). So, we need to consider basis elements of coordinate spaces

mapping to basis elements of coordinate spaces through some sort of matrix multiply. Let's first lay out our coordinate spaces. We want \mathbb{R}^n to be the coordinate space for V according to a given basis. This means that we want everything in \mathbb{R}^n to act just like vectors in V , only just look different. So we want $T_1 : V \rightarrow \mathbb{R}^n$ to be an isomorphism. Similarly, we want $T_2 : \mathbb{R}^m \rightarrow W$ to be an isomorphism. Finally, M will somehow have encoded in it, the linear transformation working as a matrix multiply. This all seems pretty far out there, but it will work. Don't just trust me though! This is mathematics, let's see why this will work. First, we should look at an outline on how we want this to work.

Outline:

- Given two vector spaces V and W .
- Considering a transformation $T : V \rightarrow W$.
- Choose a basis $\mathcal{B}_V = \{v_1, v_2, \dots, v_n\}$ of V .
- Choose a basis $\mathcal{B}_W = \{w_1, w_2, \dots, w_m\}$ of W .
- Want M so that $[T(v)]_{\mathcal{B}_W} = M[v]_{\mathcal{B}_V}$.

Notice that we will work with vectors in the coordinate spaces and interpret results afterward. **Notice also, that all of this depends strictly on the bases chosen.** This is important. So, here's how we get M . On the left in Figure 15 we see what some linear transformation T might do with basis elements of V . Notice that $\text{ran}(T)$ in this case is not W , but a subspace of W . Notice also that $\text{null}(T) \neq \{0\}$. M should do the same thing to the coordinate spaces (See the right figure in Figure 15).

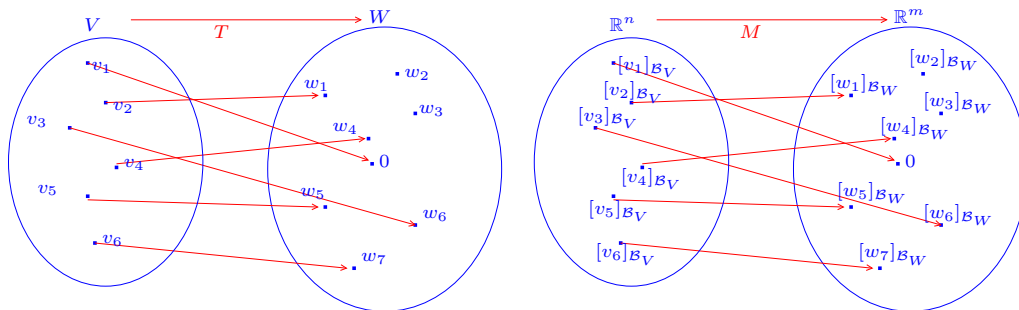


Figure 15:

We then need to create a matrix that does what we want: $[T(v)]_{\mathcal{B}_W} = M[v]_{\mathcal{B}_V}$. We do this one basis element at a time. We want to construct the matrix so that

$$\begin{aligned}
 M[v_1]_{\mathcal{B}_V} &= [T(v_1)]_{\mathcal{B}_W} \\
 T[v_2]_{\mathcal{B}_V} &= [T(v_2)]_{\mathcal{B}_W} \\
 &\vdots \\
 T[v_n]_{\mathcal{B}_V} &= [T(v_n)]_{\mathcal{B}_W}.
 \end{aligned}$$

Notice that

$$\begin{aligned} [v_1]_{\mathcal{B}_V} &= \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ [v_2]_{\mathcal{B}_V} &= \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ &\vdots \end{aligned}$$

So, $[T(v_1)]_{\mathcal{B}_W} = M[v_1]_{\mathcal{B}_V}$ gives the first column of M , $[T(v_2)]_{\mathcal{B}_W} = M[v_2]_{\mathcal{B}_V}$ gives the second column of M , \dots , and $M[T(v_n)]_{\mathcal{B}_W} = [v_n]_{\mathcal{B}_V}$ gives the n th column of M . Thus, we have that

$$M = \begin{pmatrix} | & | & & | \\ [T(v_1)]_{\mathcal{B}_W} & [T(v_2)]_{\mathcal{B}_W} & \dots & [T(v_n)]_{\mathcal{B}_W} \\ | & | & & | \end{pmatrix}$$

Let's see this with an example.

Example 14.2. Let $V = \{ax^2 + (b-a)x + (a+b) \mid a, b \in \mathbb{R}\}$ and let $W = \mathcal{M}_{2 \times 2}$. Consider the transformation $T : V \rightarrow W$ defined by

$$T(ax^2 + (b-a)x + (a+b)) = \begin{pmatrix} a & b-a \\ a+b & a+2b \end{pmatrix}.$$

We can show that T is linear (be sure you know how to do this). So, we can find a matrix representation, M , of T . First, we must find bases for V and W so that we can consider the coordinate spaces and figure out how big M must be.

$$V = \{ax^2 + (b-a)x + (a+b) \mid a, b \in \mathbb{R}\} = \text{span} \{x^2 - x + 1, x + 1\}.$$

So a basis for V is $\mathcal{B}_V = \{x^2 - x + 1, x + 1\}$. We can use the standard basis for $\mathcal{M}_{2 \times 2}$. Since V is a 2-dimensional space, its coordinate space is \mathbb{R}^2 . W being a 4-dimensional space means that its coordinate space is \mathbb{R}^4 . So, we need to find M that will multiply by a vector in \mathbb{R}^2 and output a vector in \mathbb{R}^4 . This means that M must be a 4×2 matrix. We also want M to act like T . That is, we want $[T(v)]_{\mathcal{B}_W} = M[v]_{\mathcal{B}_V}$. So, we need to see where the basis elements of V get mapped.

$$\begin{aligned} T(x^2 - x + 1) &= \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \\ T(x + 1) &= \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix} \end{aligned}$$

Now, we need to write these as coordinate vectors in \mathbb{R}^4 according to the standard basis for $\mathcal{M}_{2 \times 2}$. That is,

$$[T(x^2 - x + 1)]_{\mathcal{B}_W} = \left[\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \right]_{\mathcal{B}_W} = \begin{pmatrix} 1 \\ -1 \\ 1 \\ 1 \end{pmatrix}$$

$$[T(x + 1)]_{\mathcal{B}_W} = \left[\begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix} \right]_{\mathcal{B}_W} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix}$$

Thus,

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

We should check this to be sure we agree that this matrix multiply will do what we want. Let's pick the vector $v = 2x^2 + 2x + 6$. I know that $v \in V$ because I just chose $a = 2, b = 4$ and put them into $ax^2 + (b - a)x + (a + b)$. Anyway, according to the definition for T , we get

$$T(v) = T(2x^2 + (4 - 2)x + (2 + 4)) = \begin{pmatrix} 2 & 4 - 2 \\ 2 + 4 & 2 + 2(4) \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 6 & 10 \end{pmatrix}.$$

Now, we try the matrix multiply. The way I chose v tells us that $v = 2(x^2 - x + 1) + 4(x + 1)$. So

$$[v]_{\mathcal{B}_V} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}.$$

Now, we compute

$$[T(v)]_{\mathcal{B}_W} = M[v]_{\mathcal{B}_V} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \cdot 2 + 0 \cdot 4 \\ -1 \cdot 2 + 1 \cdot 4 \\ 1 \cdot 2 + 1 \cdot 4 \\ 1 \cdot 2 + 2 \cdot 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 6 \\ 10 \end{pmatrix}.$$

This is exactly what we wanted! I'll show you why... Earlier we found that

$$T(v) = \begin{pmatrix} 2 & 2 \\ 6 & 10 \end{pmatrix}.$$

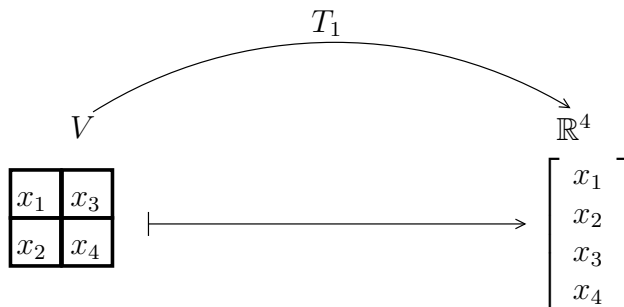
Let's write this as a coordinate vector. But, first notice that

$$\begin{pmatrix} 2 & 2 \\ 6 & 10 \end{pmatrix} = 2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + 2 \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + 6 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + 10 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

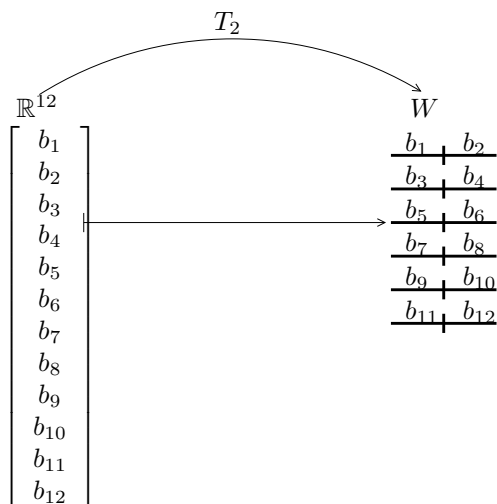
So,

$$\left[\begin{pmatrix} 2 & 2 \\ 6 & 10 \end{pmatrix} \right]_{\mathcal{B}_W} = \begin{pmatrix} 2 \\ 2 \\ 6 \\ 10 \end{pmatrix}$$

When working with radiographic transformations, we found a matrix using tomomap. But, our objects weren't vectors in \mathbb{R}^N that could be multiplied by the matrix we found. Let's use the above information to explore, through an example, what was really happening. Let $V = \mathcal{I}_{2 \times 2}$, the space of 2×2 objects. Let T be the radiographic transformation with 6 views having 2 pixels each. This means that the codomain is the set of radiographs with 12 pixels. To figure out the matrix M representing this radiographic transformation, we first change the objects in V to coordinate vectors in \mathbb{R}^4 via the isomorphism T_1 . So T_1 is defined by:



After performing the matrix multiply, we will change from coordinate vectors in \mathbb{R}^{12} back to radiographs via T_2 . So T_2 is defined by:



Our radiographic transformation is then represented by the matrix M (which we called T in the labs). M will be a 12×4 matrix determined by the radiographic set up. We've computed M several times in previous labs, but the real mathematics was all a bit hand-wavy and so now we see that really, what we have is that T maps from V to W by taking a side excursion through coordinate spaces and doing a matrix multiply.

Note: Typically, to simplify notation, we write $T(v) = Mv$ when we actually mean $[T(v)]_{\mathcal{B}_W} = M[v]_{\mathcal{B}_V}$. This is understandable by a mathematician because we recognize that when two spaces are isomorphic, they are "essentially" the same set. The only difference is that the vectors look different. In this class, we will maintain the notation discussed in this section being aware that this is just to get used to the ideas before relaxing our notation.

Exercises

Find the matrix representation of the transformation $T : V \rightarrow W$ in each of the exercises below.

1. $V = \mathcal{P}_3$ with the basis $\mathcal{B} = \{x^3, x^2 + 1, x + 1, 1\}$ and $W = \mathcal{P}_2$ with the standard basis.
 $T(ax^3 + bx^2 + cx + d) = 3ax^2 + 2bx + c$.

2. $V = \mathbb{R}^3$ with the standard basis and $W = \mathbb{R}^3$ with the basis $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$.

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

3. $V = \mathcal{M}_{4 \times 3}$ with the standard basis and $W = \mathcal{M}_{2 \times 2}$ with the basis

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right\}.$$

$$T \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{pmatrix} = \begin{pmatrix} a_{11} + a_{12} + a_{13} & a_{21} + a_{22} + a_{23} \\ a_{31} + a_{32} + a_{33} & a_{41} + a_{42} + a_{43} \end{pmatrix}$$

15 Matrix Spaces

In this section we will use the idea that any linear transformation has a matrix representation. We will define spaces similar to the spaces for linear transformations. So, let's start with a reminder of some definitions.

Definition 15.1. Let V and W be vector spaces and let $T : V \rightarrow W$ be a linear transformation. We then define the following spaces corresponding to T .

The nullspace of T , is the subspace of V that map to $0 \in W$. That is,

$$\text{null}(T) = \{v \in V \mid T(v) = 0\}.$$

We say that the nullity of T , is the dimension of the subspace $\text{null}(T)$.

We say that the codomain of T , is the ambient vector space to which we map. That is, the codomain of T is W .

We say that the range of T , is the subspace of the codomain W that contains all of the outputs from V under the transformation T . That is,

$$\text{ran}(T) = \{T(v) \mid v \in V\}.$$

We say that the rank of T , is the dimension of the subspace $\text{ran}(T)$.

Note, also, that $\text{nullity}(T) \leq \dim V$ since $\text{null}(T)$ is a subspace of V . And $\text{rank}(T) \leq \dim W$ since $\text{ran}(T)$ is a subspace of W .

15.1 The Nullspace of a Matrix

To find the nullspace, $\text{null}(T)$, for some transformation $T : V \rightarrow W$, we go straight to the definition. We are looking for all vectors $v \in V$ so that

$$T(v) = 0.$$

Now, suppose that M is the matrix representation of T when working with the basis, \mathcal{B}_V , of V . Then this equation becomes

$$M[v]_{\mathcal{B}_V} = 0.$$

We can solve for v using our most basic of linear algebra tools: matrix reduction! So, this may be tedious, but we can do it.

The take-away here is that the nullspace, $\text{null}(M)$, of an $m \times n$ matrix M is really the nullspace of the corresponding transformation, T , where $T(x) = Mx$ for all $x \in \mathbb{R}^n$. That is, $\text{null}M = \{x \in \mathbb{R}^n \mid T(x) = Mx = 0\}$. The term nullity carries over to matrices as well. We define the nullity of a matrix to be the dimension of the nullspace. Let's try a couple examples.

Example 15.1. *Given the matrix*

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

We can find $\text{null}(M)$. This means that we need to find $\text{nulls}(T)$ where T is defined by $Tx = Mx$. To do this, we solve the equation

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

We do this using matrix reduction.

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ -1 & 0 & 2 \end{pmatrix} \xrightarrow[\substack{R_2 = -2r_1 + r_2 \\ R_3 = r_1 + r_3}]{R_1 = r_2 + r_1, R_2 = -r_2} \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & -3 \\ 0 & 1 & 3 \end{pmatrix} \xrightarrow[\substack{R_3 = r_2 + r_3}]{R_1 = r_2 + r_1, R_2 = -r_2} \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}.$$

Thus

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2z \\ -3z \\ z \end{pmatrix}.$$

And we now find that

$$\text{null}(M) = \text{span} \left\{ \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} \right\}.$$

Example 15.2. *Let $V = \{ax^3 + bx^2 - ax + c \mid a, b, c \in \mathbb{R}\}$ and $W = \mathcal{M}_{2 \times 3}$. Now, let us consider the transformation $T : V \rightarrow W$ defined by*

$$T(ax^3 + bx^2 - ax + c) = \begin{pmatrix} a & a & a \\ a + b & -a & -b \end{pmatrix}.$$

We want to find the matrix representation, M , of T and then find $\text{null}(M)$. First, we find a basis for V .

$$V = \{ax^3 + bx^2 - ax + c \mid a, b, c \in \mathbb{R}\} = \text{span} \{x^3 - x, x^2, 1\}.$$

Thus, a basis for V is

$$\mathcal{B}_V = \{x^3 - x, x^2, 1\}.$$

Now we need to find where each basis element maps and to write them as coordinate vectors (We'll use the standard basis for W).

$$[T(x^3 - x)]_{\mathcal{B}_W} = \left[\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} \right]_{\mathcal{B}_W} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ 0 \end{pmatrix}$$

$$[T(x^2)]_{\mathcal{B}_W} = \left[\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix} \right]_{\mathcal{B}_W} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}$$

$$[T(1)]_{\mathcal{B}_W} = \left[\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right]_{\mathcal{B}_W} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Thus, the matrix representation is

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

Now, to find $\text{null}(M)$, we solve the equation

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} [v]_{\mathcal{B}_V} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

We will do this using matrix reduction (I will leave out the directions as they can be tedious on this example.):

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{array} \right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{array} \right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right).$$

This gives us the solution

$$[v]_{\mathcal{B}_V} = \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix}, \text{ where } z \text{ can be any real number.}$$

That is,

$$\text{null}(M) = \text{span} \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

15.2 The Columns Space of a Matrix

As with the nullspace, we can define the range space of a transformation with a matrix representation M which is an $m \times n$ matrix. Here, we are discussing the range space of the linear transformation T , defined by $T(x) = Mx$ for all $x \in \mathbb{R}^n$. This space is actually called the *column space* of the matrix M . Let's see why. First recall that $\text{ran}(T) = \{T(x) \in \mathbb{R}^m \mid x \in \mathbb{R}^n\} = \{Mx \in \mathbb{R}^m \mid x \in \mathbb{R}^n\}$. Let's write this out. Let M be the matrix with columns c_1, c_2, \dots, c_n . That is,

$$M = \left(\begin{array}{c|c|c|c} | & | & & | \\ c_1 & c_2 & \dots & c_n \\ | & | & & | \end{array} \right).$$

If T is as we described above then

$$\begin{aligned} \text{ran}(T) &= \{Mx \in \mathbb{R}^m \mid x \in \mathbb{R}^n\} \\ &= \left\{ \left(\begin{array}{c|c|c|c} | & | & & | \\ c_1 & c_2 & \dots & c_n \\ | & | & & | \end{array} \right) \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{R}^m \mid \alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R} \right\} \\ &= \{\alpha_1 c_1 + \alpha_2 c_2 + \dots + \alpha_n c_n \mid \alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}\}. \end{aligned}$$

Thus, $\text{ran}(T)$ is the span of the columns of M . So, we call $\text{ran}(T)$ the column space of M and write $\text{col}(M)$. Note: This does not mean that the columns form a basis for $\text{col}(M)$. Let's look further into this.

But first note that the term rank carries over to matrices. So, we define the rank of a matrix to be the dimension of the column space.

Let's look at another way to find $\text{col}(M)$. This will help us find a basis. We want to find all $w \in \mathbb{R}^m$ so that there is an $v \in \mathbb{R}^n$ so that $Mv = w$. Again, matrix reduction is very useful here. Let us look at some examples.

Example 15.3. Let us consider the matrix

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

To find $\text{col}(M)$, we set up the equation $Mv = b$ and determine all possible b .

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ -1 & 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

Using matrix reduction we get

$$\begin{aligned} & \left(\begin{array}{ccc|c} 1 & 1 & 1 & a \\ 2 & 1 & -1 & b \\ -1 & 0 & 2 & c \end{array} \right) \xrightarrow[R_3=r_1+r_3]{R_2=-2r_1+r_2} \left(\begin{array}{ccc|c} 1 & 1 & 1 & a \\ 0 & -1 & -3 & -2a+b \\ 0 & 1 & 3 & a+c \end{array} \right) \\ & \xrightarrow[R_3=r_2+r_3]{R_1=r_2+r_1, R_2=-r_2} \left(\begin{array}{ccc|c} 1 & 0 & -2 & -a+b \\ 0 & 1 & 3 & 2a-b \\ 0 & 0 & 0 & -a+b+c \end{array} \right). \end{aligned}$$

Notice that as long as $w = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ with $-a + b + c = 0$ the equation $Mv = w$ has a solution.

Thus,

$$\begin{aligned} \text{col}(M) &= \left\{ \left(\begin{array}{c} a \\ b \\ c \end{array} \right) \mid -a + b + c = 0 \right\} = \left\{ \left(\begin{array}{c} b+c \\ b \\ c \end{array} \right) \mid b, c \in \mathbb{R} \right\} \\ &= \text{span} \left\{ \left(\begin{array}{c} 1 \\ 1 \\ 0 \end{array} \right), \left(\begin{array}{c} 1 \\ 0 \\ 1 \end{array} \right) \right\} \end{aligned}$$

Notice that one of the basis elements we found corresponds to one of the columns of M (the second column of M to be exact). In fact, one of the leading coefficients is in the second column of M and so therefore corresponds to that column. It seems that it would be very nice to not have this extra step at the end after reducing the matrix. We see that there are two leading ones and there are two basis elements. We also see that the columns in M corresponding to the leading ones are linearly independent. It turns out that we can actually pick out these columns (corresponding to the columns with leading entries after being reduced) to make a basis. Let's see why.

Notice that if $\text{col}(M) = \text{span} \{c_1, c_2, \dots, c_n\}$, then to form a basis, we need only find the columns that are linearly independent. Here's a process to see that grabbing the columns corresponding to the leading ones works.

Let's choose c_1 to be the first basis element for $\text{col}(M)$. (Note: if c_1 is all zeros, we just start with the first column that isn't all zeros.) Since c_1 is not the zero vector, $\{c_1\}$ is linearly independent. Now, we check to see if $\{c_1, c_2\}$ is linearly independent. We can do this by solving for α in the equation $\alpha c_1 = c_2$. We can do this by reducing the augmented matrix

$$\left(\begin{array}{c|c} | & | \\ c_1 & c_2 \\ | & | \end{array} \right).$$

If the second column has a leading one, then that means there is a row with zeros to the left of the augment and a nonzero on the right of the augment. This would mean that the equation $\alpha c_1 = c_2$ has no solution and they are linearly independent. If there is no leading entry in the second column, then these columns are linearly dependent.

Now, we check to see if c_3 is linearly independent with c_1 and c_2 . That means we want to solve for α and β in the equation $\alpha c_1 + \beta c_2 = c_3$. This can be done by reducing the augmented matrix

$$\left(\begin{array}{cc|c} | & | & | \\ c_1 & c_2 & c_3 \\ | & | & | \end{array} \right).$$

If, after reducing, the third column has a leading entry, then $\{c_1, c_3\}$ is linearly independent and $\{c_2, c_3\}$ is also linearly independent. If not, then either $\{c_1, c_3\}$ or $\{c_2, c_3\}$ is linearly dependent. We can continue this process and see that set of columns corresponding to a leading entry in the reduced matrix is a linearly independent set. So we choose them to be in the basis for $\text{col}(M)$. All other columns are in the span of these chosen vectors.

Example 15.4. Let $V = \{ax^3 + bx^2 - ax + c \mid a, b, c \in \mathbb{R}\}$ and $W = \mathcal{M}_{2 \times 3}$. Now, let us consider the transformation $T : V \rightarrow W$ defined by

$$T(ax^3 + bx^2 - ax + c) = \begin{pmatrix} a & a & a \\ a + b & -a & -b \end{pmatrix}.$$

Recall that

$$\mathcal{B}_V = \{x^3 - x, x^2, 1\}$$

is a basis of V . Above, we found the matrix representation to be

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

To find $\text{col}(M)$, we find all $w \in \mathbb{R}^6$ so that there is a $v \in \mathbb{R}^3$ with $Mv = w$. In other words, we find all $w \in \mathbb{R}^6$ so that

$$w \in \text{span} \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

In this example, it is clear that the last column is not part of a linearly independent set. Also, it is clear that the first two columns are linearly independent (they are not multiples of

one another). Thus, a basis for the column space is

$$\left\{ \left(\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ 0 \end{array} \right), \left(\begin{array}{c} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ -1 \end{array} \right) \right\}.$$

Exercises

1. For each of the following matrices, find $\text{null}(M)$, $\text{col}(M)$, $\text{rank}(M)$, $\text{nullity}(M)$, $\text{size}(M)$, the number of columns without leading entries, and the number of leading entries in the echelon form.

(a)

$$M = \begin{pmatrix} 1 & 2 & 3 & -1 \\ 1 & 1 & -1 & -1 \\ 2 & 3 & 2 & -2 \\ 5 & 6 & -1 & -5 \end{pmatrix}$$

(b)

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

(c)

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

(d)

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

2. For each of the examples in this section, state $\text{rank}M$, $\text{nullity}(M)$, $\text{size} M$, the number of columns without leading entries, and the number of leading entries in the echelon form. (Note: you already know $\text{null}(M)$ and $\text{col}(M)$ for each of these matrices.)
3. How does $\text{nullity}(M)$ show up in the echelon form of the matrix reduction?
4. How does $\text{rank}(M)$ show up in the echelon form of the matrix reduction?
5. How are $\dim V$ and $\dim W$ related to M ?
6. Use the Rank-Nullity Theorem to make a conjecture that brings together a relationship with all or some of these.

15.3 Onto, One-to-One, and what this means about Matrices

This section will bring together the properties of a linear transformation with properties about the matrix representation. Mostly, I'm going to ask you some questions.

First, let's recall what we know about one-to-one and onto transformations. In Theorem 13.3, we found that another definition for an onto transformation, $T : V \rightarrow W$ is that $\text{ran}(T) = W$. In Theorem 13.4, we found that another definition for a 1-1 transformation $T : V \rightarrow W$ is that $\text{null}(T) = \{0\}$. Use these and what you found in the previous section to answer the following questions.

Exercises

1. If $T : V \rightarrow W$ and M is the matrix representation of T . What can you say about the size of M ?
2. If $T : V \rightarrow W$ and M is the $n \times m$ matrix representation of T . What can you say about n and m if T is onto?
3. If $T : V \rightarrow W$ and M is the $n \times m$ matrix representation of T . What can you say about n and m if T is one-to-one?
4. If $T : V \rightarrow W$ and M is the $n \times m$ matrix representation of T . What can you say about n and m if T is a bijection?
5. Let T be the radiographic transformation described by the following scenario.
 - Height and width of image in voxels: $n = 2$ (Total voxels $N = 4$)
 - Pixels per view in radiograph: $m = 2$
 - *ScaleFac* = 1
 - Number of views: $a = 3$
 - Angle of the views: $\theta_1 = 0^\circ$, $\theta_2 = 45^\circ$, and $\theta_3 = 90^\circ$

For ease of notation, call the corresponding matrix T . Determine whether T is one to one, onto, both, or neither. You may use `tomomap` to find T .

6. For each of the transformations in Exercise 1 in Section 10, find the associated matrix.
7. For each of the transformations:
 - A. T_3 the third radiographic transformation from Lab 2.-
 - B. $\frac{d}{dx} : \mathcal{P}_4 \rightarrow \mathcal{P}_3$
 - C. $R_\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, where θ is a specific fixed angle measured from the x axis and R_θ is the rotation transformation
 - D. $\pi_v : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, the component of v parallel to the x axis.
 - E. $D : \mathbb{R}^3 \rightarrow \mathbb{R}$ defined by $D(v) = v \cdot u$, where $u = (0, 1, -1)$.

Answer each of the following:

- (a) What is $\text{null}(T)$?
- (b) What is $\text{ran}(T)$?
- (c) Is the transformation 1-1? Prove it.
- (d) Is the transformation onto? Prove it.

16 Inverse Transformations

You might wonder why we would want to use matrices at all. After all, we already have a way of computing the outputs of a transformation. What would getting a matrix do to help us? Even algorithms can be computed to find these outputs without having a matrix representation. So, all this complicated stuff gets in the way, right? Well...there's so much more to Linear Algebra than

- Determining if something is a vector space.
- Determining if something is in a vector space.
- Determining linear dependence.
- Finding the span of vectors.
- Computing the outputs of linear transformations.
- Solving systems of equations.
- \vdots

16.1 A (Very) Brief Introduction to Tomography

In this class, we've actually talked about an application: Radiography/Tomography. Well, we haven't talked about the Tomography part. We will get there. First note that nobody ever really computes the radiograph. This is done using a machine that sends x-rays through something. But what people want to be able to do is to figure out what the object being radiographed looks like. This is the idea behind Tomography. So, we do need to be able to find the object that was produced by a certain transformation.

Suppose we know that $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the transformation given by

$$T(v) = Mv.$$

where $M \in \mathcal{M}_{m \times n}$. Suppose also that we know that the vector $w \in \mathbb{R}^m$ is obtained by Tv for some v , but we don't know which v . How would we find it? Let's look at some examples.

Example 16.1. Suppose we want to know what $v \in V$ produces $w = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ in the equation

$Tv = Mv = w$ where

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

We can find v so that $Tv = w$ by solving the equation:

$$Tv = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

which is the same as

$$Mv = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

This leads to the system of equations

$$\begin{aligned} x_1 + x_2 &= 1 \\ x_1 + 2x_2 &= 1 \end{aligned}$$

We use matrix reduction to solve this system.

$$\left(\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 2 & 1 \end{array} \right) \rightarrow \left(\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 1 & 0 \end{array} \right) \rightarrow \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 0 \end{array} \right) \Rightarrow x_1 = 1, x_2 = 0 \Rightarrow v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Example 16.2. Let $m = 2, n = 4, w = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and

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

We can find v so that $Tv = w$ by solving the equation:

$$Tv = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

$$Mv = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 1 & 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

Which leads to the system of equations

$$\begin{aligned} x_1 + x_2 + 3x_4 &= 1 \\ x_1 + 2x_2 + x_3 + x_4 &= 3 \end{aligned}$$

We can use a matrix to solve this system, reduce it and find v :

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

This gives us that

$$x_1 = x_3 - 5x_4 - 1 \quad \text{and} \quad x_2 = 2x_4 - x_3 + 2.$$

So, there are infinitely many v that could have led us to the result w through this transformation. One particular v is found when we let $x_3 = x_4 = 1$. In this case, we get

$v = \begin{pmatrix} -5 \\ 3 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ The next step would be to figure out the best of these to choose (but that's for later).

Example 16.3. Let $m = 3$, $n = 2$, $w = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and

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

Again, we can set up the equation $Tv = w$ and solve for v .

$$\begin{aligned} Tv &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ Mv &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \end{aligned}$$

Which leads to the system of equations

$$\begin{aligned} x_1 &= 1 \\ 2x_1 + x_2 &= 1 \\ x_1 + 2x_2 &= 1 \end{aligned}$$

We see that $x_1 = 1$, but then $x_2 = -1$ and $x_2 = 0$. This gives us that there is no v that would produce w under this transformation.

In each of the above examples, we may want to find each of the elements of the domain that would produce several elements in the codomain: $T(v_1) = w_1, T(v_2) = w_2, \dots, T(v_k) = w_k$. This would take quite a long time to compute these matrix reductions every time (granted there are computers to do this, but in real life the matrices you use are MUCH bigger). So, we need another way.

16.2 Finding solutions to $Mv = w$.

Notice that if we were to change the right-hand-side in an example above, the matrix reduction steps would not change. So, if we wanted to do this for many different outputs, we could set up the matrix like:

$$\left(M \left| \begin{array}{c|c|c|c} | & | & \dots & | \\ w_1 & w_2 & \dots & w_k \\ | & | & & | \end{array} \right. \right). \quad (5)$$

Where this augmented matrix has the matrix M on the left side of the augment and column vectors w_1, w_2, \dots, w_k on the right side of the augment. We could then perform the matrix reduction all at once.

Example 16.4. Find v_1, v_2, v_3 when $w_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $w_2 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, and $w_3 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and

$$M = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

by solving $Mv_1 = w_1$, $Mv_2 = w_2$, and $Mv_3 = w_3$ all at once using an augmented matrix of the form in (5).

We set up the augmented matrix

$$\left(\begin{array}{cc|ccc} 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 3 & -1 \end{array} \right) \rightarrow \left(\begin{array}{cc|ccc} 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & 4 & 0 \end{array} \right) \rightarrow \left(\begin{array}{cc|ccc} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 & 0 \end{array} \right) \rightarrow \left(\begin{array}{cc|ccc} 1 & 0 & 0 & -1 & 1 \\ 0 & 1 & 1 & 2 & 0 \end{array} \right).$$

Using the above, we get that the solution to $Mv_1 = w_1$ is $v_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, to $Mv_2 = w_2$ is $v_2 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$, and to $Mv_3 = w_3$ is $v_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

Still, if there are a large number of w_i , then we would still have a lot of work to do.

But, we know that all the w_i come from the vector space \mathbb{R}^m . This means that they can each be written as a linear combination of e_1, e_2, \dots, e_m , the standard basis elements. So, let's look at what we can do to simplify all of this.

Exercises

- Knowing that T is a linear transformation and $w = \alpha_1 e_1 + \alpha_2 e_2 + \dots + \alpha_m e_m$, what can you say about where the basis vectors $v_1, v_2, \dots, v_m \in \mathbb{R}^m$ map in each of the following cases?
 - T is onto.
 - T is one-to-one.
 - T is a bijection.

You may want to review some of your answers in the previous section about these cases.

- Knowing that T is linear, how can the above answers tell us which v gives us $T(v) = w$ when T is a bijection?
- Suppose T isn't a bijection now. Recall that this means that $\text{null}(T)$ contains a basis from $V = \mathbb{R}^m$. What does this mean when we look for v that gives us $Tv = w$ for some specific w ?

16.3 Finding solutions to $Mv = e_k$, an element of the standard basis.

All your answers should lead us to the fact that, if T is a bijection, we can see what maps to each of the standard basis elements in \mathbb{R}^m and be able to state where each vector w came from. Let's look at an example.

Example 16.5. Let $m = n = 2$, $M = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. Suppose $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is defined by $T(v) = Mv$. We will use the standard basis for \mathbb{R}^2 . We want to know what is mapped to $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and what is mapped to $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ so we set up and reduce the augmented matrix representing $Mv_1 = e_1$ and $Mv_2 = e_2$. Notice, this can be done with one augmented matrix.

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

Notice that this example tells us that if we are solving $T(v_1) = e_1$ (that is, $Mv_1 = e_1$) then $v_1 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$ and if we are solving $T(v_2) = e_2$ (that is, $Mv_2 = e_2$) then $v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Example 16.6. Now, to get back to a particular w . For $w = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ say we want to find v so that $T(v) = w$. Using $T(v) = Mv$ we can solve $Mv = w$.

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

So, $v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ maps to w

We can now write v as a linear combination of the vectors v_1 and v_2 found above.

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 2 \\ -1 \end{pmatrix} + 2 \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Example 16.7. Let's try another w . For $w = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$ we want to find v that gives us $Mv = w$.

$$\left(\begin{array}{cc|c} 1 & 1 & 3 \\ 1 & 2 & 5 \end{array} \right) \rightarrow \left(\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & 1 & 2 \end{array} \right) \rightarrow \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array} \right).$$

So, $v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ maps to w . Again, we write v as a linear combination of v_1 and v_2 .

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} = 3 \cdot \begin{pmatrix} 2 \\ -1 \end{pmatrix} + 5 \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

More Exercises

- Use the matrix M above. For $w = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$ do each of the following.
 - Find v that gives us $Mv = w$.
 - Write v as a linear combination of v_1 and v_2 in Example 16.5.
- Now make a conjecture about how you can find v if you know how to get e_1 and e_2 through this transformation. Solidify your conjecture by writing down a matrix multiply that you could use to get v . Hint: this matrix is "right" before your eyes.

16.4 The Inverse Matrix, when $m = n$

The matrix that you found before is an important matrix. We call this matrix the inverse matrix. The process that we used is how you would find the inverse to any square matrix. Let us define the inverse more rigorously.

Definition 16.1. Let M be an $n \times n$ square matrix. We say that M has an inverse (and call it M^{-1}) if and only if

$$MM^{-1} = M^{-1}M = I_{n \times n},$$

where $I_{n \times n}$ is the square identity matrix. If M has an inverse, we say that it is invertible.

We will now use the process above to find the inverse of some matrices.

Example 16.8. We want to find M^{-1} when

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

Following the procedure above with an augmented matrix we find the inverse matrix M^{-1} . That is, we reduce the matrix $(M \mid e_1 \ e_2)$ as follows.

$$\left(\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array} \right) \rightarrow \left(\begin{array}{cc|cc} 1 & 0 & 1 & -3 \\ 0 & 1 & 0 & 1 \end{array} \right)$$

We see that

$$M \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad M \begin{pmatrix} -3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

So we know that

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

Thus,

$$M^{-1} = \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix}.$$

We've seen examples where the elimination process does not give us a leading 1 in every row. This next example shows us what this means.

Example 16.9. $M = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$

Again, we begin by reducing the augmented matrix $(M \mid e_1 \ e_2)$ as follows:

$$\left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{array} \right) \rightarrow \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 0 & -2 & 1 \end{array} \right).$$

This cannot be reduced further. From the reduced echelon form, we see that there is no vector v so that $Mv = e_1$ nor is there a vector v so that $Mv = e_2$. That is, there is no matrix M^{-1} so that $MM^{-1} = I_{2 \times 2}$, the identity matrix. So, M does not have an inverse.

Now, let's look at a bigger example.

Example 16.10. We will find the inverse of the matrix

$$M = \begin{pmatrix} 1 & 1 & -2 \\ 1 & 2 & 1 \\ 2 & -1 & 1 \end{pmatrix}.$$

No matter the size of M , we always start by reducing the augmented matrix $(M \mid e_1 \ e_2)$.

$$\begin{aligned} \left(\begin{array}{ccc|ccc} 1 & 1 & -2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 2 & -1 & 1 & 0 & 0 & 1 \end{array} \right) &\rightarrow \left(\begin{array}{ccc|ccc} 1 & 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & -3 & 5 & -2 & 0 & 1 \end{array} \right) &\rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & -5 & 2 & -1 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 0 & 14 & -5 & 3 & 1 \end{array} \right) \\ &\rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & -5 & 2 & -1 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 0 & 1 & -5/6 & 1/2 & 1/6 \end{array} \right) &\rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 3/14 & 1/14 & 5/14 \\ 0 & 1 & 0 & 1/14 & -5/14 & -3/14 \\ 0 & 0 & 1 & -5/14 & 3/14 & 1/14 \end{array} \right). \end{aligned}$$

Using the above understanding, we see that

$$M \begin{pmatrix} 3/14 & 1/14 & 5/14 \\ 1/14 & -5/14 & -3/14 \\ -5/14 & 3/14 & 1/14 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus, we know that the inverse is

$$M^{-1} = \begin{pmatrix} 3/14 & 1/14 & 5/14 \\ 1/14 & -5/14 & -3/14 \\ -5/14 & 3/14 & 1/14 \end{pmatrix}.$$

Rest of the Exercises for Section 16

1. Find M^{-1} for

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

2. Let $M = \begin{pmatrix} 1 & 0 & -1 \\ 3 & 1 & -1 \\ -2 & -1 & 1 \end{pmatrix}$, find M^{-1} .

3. Given $M = \begin{pmatrix} 1 & 1 & -2 \\ 1 & 2 & 1 \\ 2 & 3 & -1 \end{pmatrix}$, find M^{-1} .

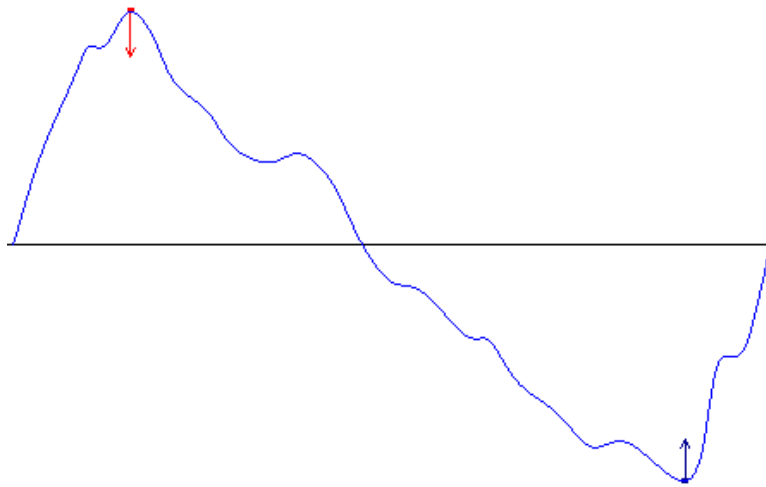
17 Heat States in a (Vector) State Space

Recall in Exercise 10 Section 4.2, we introduced the application of diffusion welding. A manufacturing company uses the process of diffusion welding to adjoin several smaller rods into a single longer rod. The diffusion welding process leaves the final rod heated to various temperatures along the rod with the ends of the rod having the same temperature. Every a cm along the rod, a machine records the temperature difference from the temperature at the ends to get heat states. We assume that the rod is thoroughly insulated except at the ends so that heat is only lost out the end.

We want to explore this application further. Let's define a function $f : [a, b] \rightarrow \mathbb{R}$ that measures, along a rod of length $L = b - a$, the difference in temperature from the temperature at the ends of the rod ($f(a) = f(b) = 0$). The quantity $f(x)$ is the temperature of the rod at position x for any x in the interval $[a, b]$.

We know that as time goes on, the heat will spread along the rod. We can imagine that after a very long time the heat will be the same all along the rod. We can even predict how the heat will diffuse:

1. Places where the temperature is higher than the surrounding area will cool.
2. Places where the temperature is cooler than the surrounding area will warm.
3. As the heat diffuses, sharpest temperature differences disappear first.
4. The long term behavior is that temperatures smooth out approaching $f(x) = 0$.



17.1 The Heat Equation

What we expect to happen is actually described mathematically by a partial differential equation. In fact, it is described by the heat equation, or diffusion equation, (with fixed unit diffusion coefficient):

$$\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial x^2}.$$

Notice that our initial intuition is seen in this equation.

1. A heat state vector f is a function of both position and time, though we often do not explicitly write $f(x, t)$ for clarity sake.
2. The rate of change of temperature at any location x is proportional to the second derivative of the temperature with respect to position.
3. Temperature is changing fastest at sharp peaks or dips in temperature where the curvature is large.
4. Regions where temperature is changing slowly are where $|\partial^2 f(x)/\partial x^2|$ is small. This occurs under two different conditions:

- (a) Where the temperature profile has nearly constant slope.
 - (b) Near inflection points of the temperature profile.
5. The sign of the second derivative ensures that the heat state function does approach the zero function.

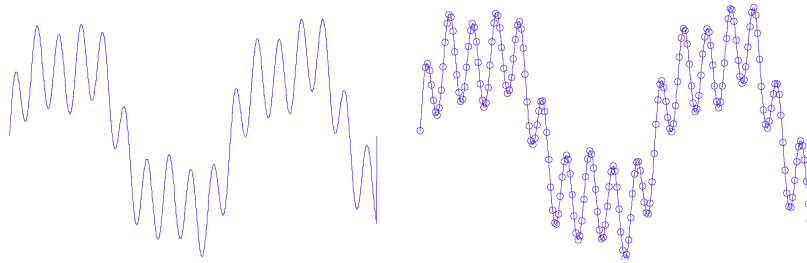
Notice that the second derivative with respect to position (x) tells us about concavity of the heat state along the bar. The “more” concave down, the faster the temperature will drop and the “more” concave up, the faster the temperature will rise. The above bulleted list just says this mathematically.

17.2 The Discrete Heat Equation, in x .

None of this seems very much like Linear Algebra. But it is! We formulate our problem in a finite dimensional vector space. We will use our knowledge of linear algebra to compute the heat state at any later time as the heat diffuses. In the previous part of the discussion we modeled the heat along a bar by a continuous function f , which we call the heat state of the bar. We will *discretize* such a heat state f by sampling the temperature at m locations along the bar. If we space the m sampling locations equally, then for $\Delta x = \frac{L}{m+1} = \frac{b-a}{m+1}$, we can choose the sampling locations to be $a + \Delta x, a + 2\Delta x, \dots, a + m\Delta x$. (We will assume that the heat state is known (and fixed) at the endpoints so we don't need to sample there!) Then, the discretized heat state is the following vector u in \mathbb{R}^{m+2} .

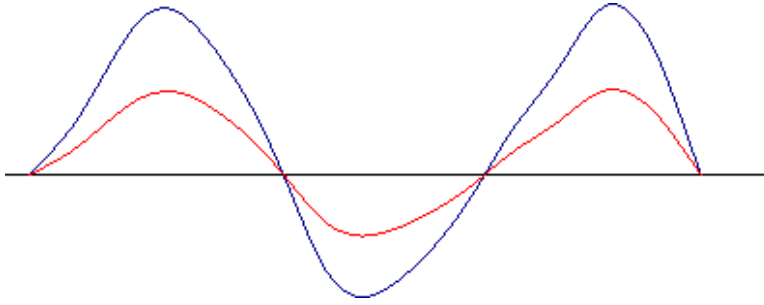
$$[0, u_1, u_2, \dots, u_m, 0] = [f(a), f(a + \Delta x), f(a + 2\Delta x), \dots, f(a + m\Delta x), f(b)].$$

Notice that $f(a) = f(b) = 0$, where $b = a + (m + 1)\Delta x$. Also, if $u_j = f(x)$ for some $x \in [a, b]$ then $u_{j+1} = f(x + \Delta x)$ and $u_{j-1} = f(x - \Delta x)$. The figure below shows a continuous heat state on the left and a heat state with sampling points marked on the right.

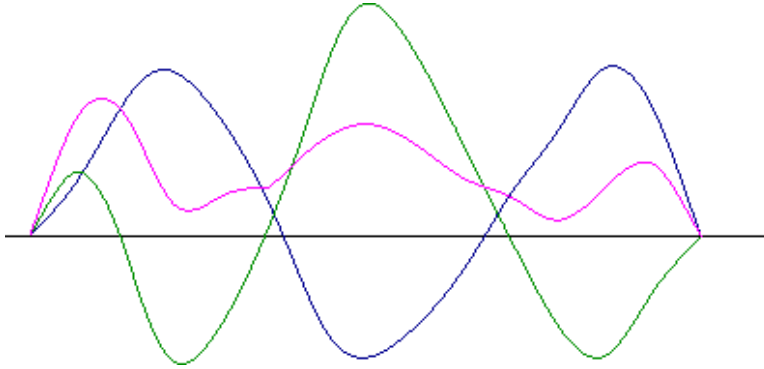


To complete our vector space development (we need the set of heat states to form a vector space so that we can use Linear Algebra in this setting), we must define heat state addition and scalar multiplication.

1. We define scalar multiplication in the usual componentwise fashion. Scalar multiplication results in a change in amplitude only.



2. We define vector addition in the usual componentwise fashion. Addition can result in changes in both amplitude and shape.



Recall, heat diffuses in the way described by

$$\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial x^2}.$$

Now, because derivatives are linear, intuitively we believe that it should be possible to get a good discrete approximation. So, we discretize the derivatives, using the definition of derivative

$$\frac{\partial f}{\partial x} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

Notice that in a discrete setting, it is not possible for $h \rightarrow 0$. The smallest h can be is Δx . Let's use this to make the matrix representation for the second derivative operator, which we call D_2 . That is, $D_2 u$ approximates $\partial^2 f / \partial x^2$.

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x} \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} \left(\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \right) \\ &\approx \frac{\partial}{\partial x} \left(\frac{f(x+\Delta x) - f(x)}{\Delta x} \right) \\ &= \frac{1}{\Delta x} \lim_{h \rightarrow 0} \left[\frac{f(x+\Delta x) - f(x) - f(x+\Delta x-h) + f(x-h)}{h} \right] \\ &\approx \frac{1}{(\Delta x)^2} [f(x + \Delta x) - 2f(x) + f(x - \Delta x)] \end{aligned}$$

Notice that we have used both a forward and backward difference definition of the derivative in order to make our approximation symmetric. This helps us keep the later linear algebra manageable. In our discretization, we have

$$\frac{\partial^2 u_j}{\partial x^2} = \frac{1}{(\Delta x)^2} (u_{j+1} - 2u_j + u_{j-1}).$$

In-Class Exercise

Write the matrix representation for the second derivative operator, D_2 with respect to the standard basis, using $(D_2(u))_j = \frac{1}{(\Delta x)^2}(u_{j+1} - 2u_j + u_{j-1})$ when $m = 6$.

Answer.

$$D_2 = \frac{1}{(\Delta x)^2} \begin{bmatrix} -2 & 1 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 1 & -2 \end{bmatrix}.$$

This says that if we want to take a discrete second derivative, we do the matrix multiply D_2u .

17.3 The Discrete Heat Equation, in t

The heat state u also changes over time. So we can write $u = u(t)$, with a time dependence.

That means we have from above $\frac{\partial^2 u(t)}{\partial x^2} = D_2u(t)$.

Now, we use a similar argument to discretize $\frac{\partial u}{\partial t}$.

$$\frac{\partial u}{\partial t} \approx \frac{u(t + \Delta t) - u(t)}{\Delta t}.$$

We want to find the diffusion operator $E : \mathbb{R}^m \rightarrow \mathbb{R}^m$, for a fixed Δt , so that $u(t + \Delta t) = Eu(t)$ (That is, $u(t + \Delta t)$ is obtained via a matrix multiply). We put the above discretizations into the heat equation to obtain:

$$\frac{u(t + \Delta t) - u(t)}{\Delta t} = \frac{1}{(\Delta x)^2} D_2u(t).$$

So,

$$u(t + \Delta t) = u(t) + \frac{\Delta t}{(\Delta x)^2} D_2u(t) = \left(I + \frac{\Delta t}{(\Delta x)^2} D_2 \right) u(t).$$

Let E be defined as the matrix $I + \frac{\Delta t}{(\Delta x)^2} D_2$. Then we have $u(t + \Delta t) = Eu(t)$.

In-Class Exercise

Find the matrix representation of E .

$$E = \begin{bmatrix} 1 - 2\delta & \delta & 0 & 0 & & \\ \delta & 1 - 2\delta & \delta & 0 & \dots & \\ 0 & \delta & 1 - 2\delta & \delta & & \\ & & \vdots & & \ddots & \end{bmatrix},$$

Here we note that we need $0 < \delta \equiv \frac{\Delta t}{(\Delta x)^2} \leq \frac{1}{4}$ for computational stability, though the proof is beyond the scope of this module (reference?). Δx is fixed so we need to take small enough time steps Δt to get this inequality. Hence as we let $\Delta x \rightarrow 0$, we are also implicitly forcing $\Delta t \rightarrow 0$.

Consider the meaning of the row and column representations of E .

1. How can we interpret the values in column k of E ?

Answer. The k th column shows how the heat at time t distributes to heat at time $t + \Delta t$ at location k in the heat state. In particular, fraction $1 - 2\delta$ of the heat at location k remains at location k and fraction δ of the heat moves to each of the two nearest neighbor locations $k + 1$ and $k - 1$.

2. How can we interpret the values in row k of E ?

Answer. The k th row shows where the heat at time $t + \Delta t$ came from. In particular, fraction $1 - 2\delta$ of the heat at location k was already at location k and fraction δ came from each of the location's two nearest neighbors.

3. How can we interpret the fact that all but the first and last columns (and rows) sum to unity?

Answer. Based on the above discussion, we see that this guarantees that no heat is lost from the system *except at the end points*. Heat is redistributed (diffused) not lost.

17.4 Heat State Evolution

Now, we consider the evolution over more than one time step.

$$\begin{aligned} u(t + \Delta t) &= Eu(t) \\ u(t + 2\Delta t) &= Eu(t + \Delta t) = E(Eu(t)) = E^2u(t) \end{aligned}$$

This means k time steps later,

$$u(t + k\Delta t) = E^k u(t)$$

Exercises

In class, we found the matrix representation of the heat diffusion operator to be

$$E = \begin{bmatrix} 1 - 2\delta & \delta & 0 & 0 & & \\ \delta & 1 - 2\delta & \delta & 0 & \dots & \\ 0 & \delta & 1 - 2\delta & \delta & & \\ & & \vdots & & \ddots & \end{bmatrix}.$$

We see that if we are want to find the k th heat state, $u(k\Delta t)$ in the heat diffusion, we need to use

$$u(k\Delta t) = E^k u(0),$$

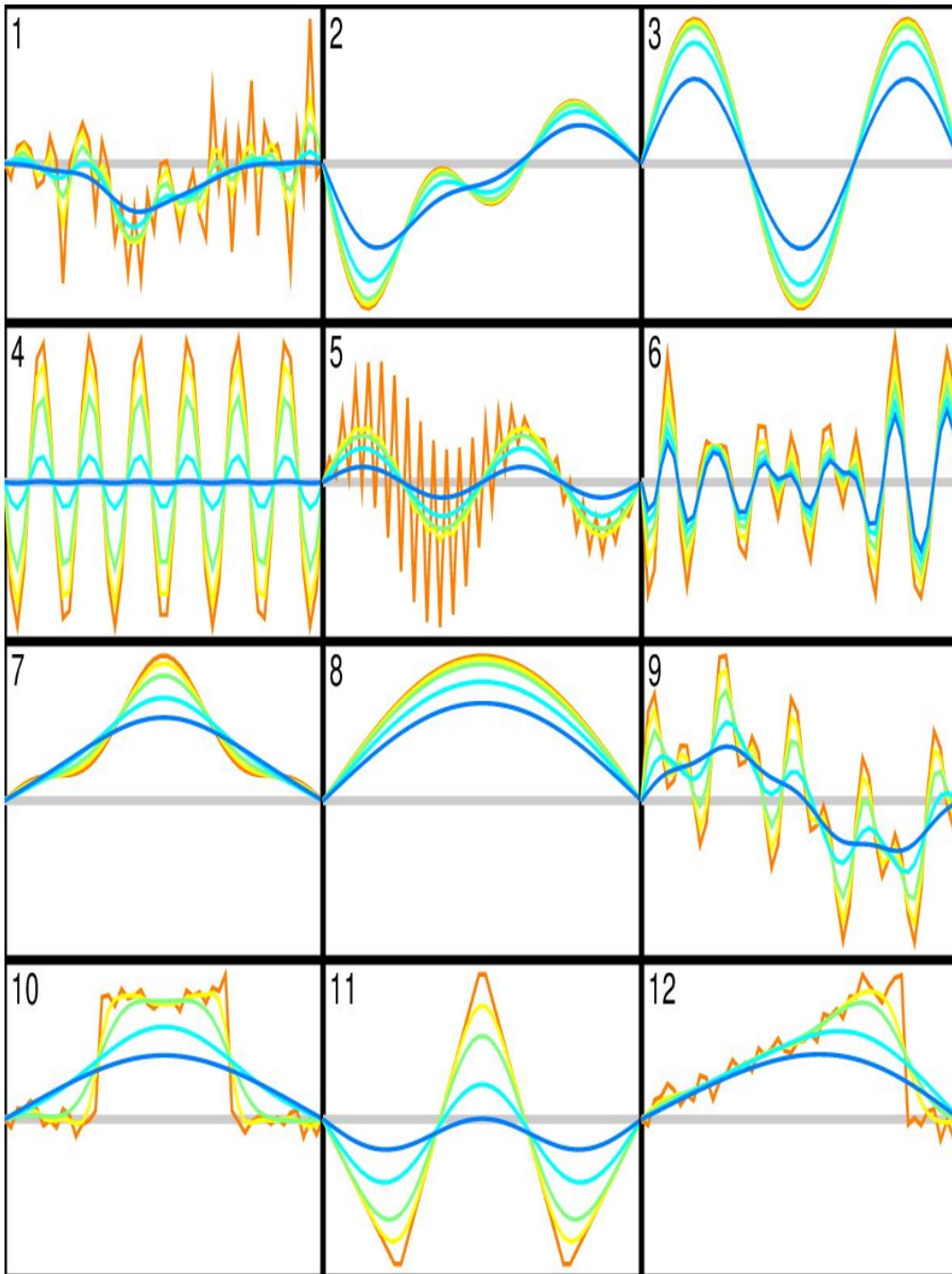
where $u(0)$ is the initial heat state. Let's explore what happens when we do these calculations.

1. What is the equation for the heat state 2 steps in the future? 1000 time steps in the future?
2. Find E^2 .

3. Does it look like raising E to higher and higher powers is an easy way to find the heat state at some time far in the future (for example, 1000 time steps away)?
4. Pick your favorite nontrivial vector $u \in \mathbb{R}^4$ and compute Eu , $E(Eu)$ and $E(E(Eu))$.
5. Does it look like multiplying over and over by the matrix E is an easy way to find the heat state at some time far in the future?
6. If you had to perform this kind of operation, what characteristics would you prefer a diffusion operator to have?

18 Heat Equation for Linear Algebra Lab 1

1. In the following picture, there are 12 different initial heat states (in orange) and their corresponding diffusions (colors varying from orange to dark blue). Group the pictures based on the diffusion behavior. Briefly list the criteria you used to group them. We will use these descriptions in the next discussion so try to be as clear as possible.
2. Write the expression for Eu for the special vectors discussed in class.
3. Now let's view the diffusion of linear combinations of these special vectors. What do you see in the diffusion of a vector that is a linear combination of
 - (a) Two of these special vectors?
 - (b) Three of these special vectors?
4. Write out algebraically what happens in the diffusion of a heat state that is a linear combination of these special vectors.
5. What if we want to find more of these special vectors? What matrix equation would we solve?
6. What do this equation and the invertible matrix theorem tell us?



19 Heat Equation for Linear Algebra Lab 2

Recall that our goal is to find the heat state after k time steps in the heat diffusion. We have observed that some heat states (simple vectors) only change in amplitude with time, $Ev_j = a_j v_j$. Now we will explore what could happen if some set of simple vectors $\beta = \{v_1, v_2, \dots, v_m\}$ is a basis for \mathbb{R}^m .

1. Working with the standard basis for \mathbb{R}^m , what can you determine about the diffusion of a heat state $u(t + \Delta t) = Eu(t)$ given basis β ?

2. Use your answer from question 1 to find a representation for the heat state after the 10th iteration of diffusion, that is, for $u(t + 10\Delta t)$.
3. Answer question 1 again, but consider a change of basis.
4. Using your answer from question 3, find a representation for the heat state after the 10th iteration of the diffusion.
5. How might these computations using the simple vector basis give us information about the long term behavior of the heat diffusion? Make some observations.

20 Eigenvalues, Eigenvectors, Eigenspaces, and Diagonalizability

In the Heat Equation Labs 1 & 2 (Sections 18 and 19), we found that some heat states will only change in amplitude in the heat diffusion (cooling) on a rod. This means that when we apply the diffusion operator to one of these heat states, we get a result that is a scalar multiple of the original heat state. Mathematically, this means

$$Ev = \lambda v, \tag{6}$$

for some scalar λ . We also saw that these vectors satisfy the matrix equation

$$(E - \lambda I)v = 0. \tag{7}$$

Since this is a homogeneous equation, we know that this equation has a solution. This means that either there is a unique solution (only the trivial solution) or infinitely many solutions. If we begin with a zero heat state (all temperatures are the same everywhere along the rod) then the diffusion is very boring (my opinion, I know) because nothing happens. It would be nice to find a nonzero vector satisfying the matrix Equation (7) because it gets us closer to the possibility of having a basis of these vectors. By the invertible matrix theorem, we know that this equation has a nonzero solution as long

$$\det(E - \lambda I) = 0.$$

Let's remind ourselves why we want such a basis. In the case of heat states, we recognize that if $\mathcal{B} = \{v_1, v_2, \dots, v_m\}$ is a basis of these special vectors so that $Ev_i = \lambda_i v_i$ and u_0 is our initial heat state, we can write u in coordinates according to \mathcal{B} . That is, there are scalars $\alpha_1, \alpha_2, \dots, \alpha_m$ so that

$$u_0 = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_m v_m.$$

Then, when we apply the diffusion operator to find the heat state, u_1 a short time later, we get

$$\begin{aligned} u_1 &= Eu_0 = E(\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_m v_m) \\ &= \alpha_1 Ev_1 + \alpha_2 Ev_2 + \dots + \alpha_m Ev_m \\ &= \alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2 + \dots + \alpha_m \lambda_m v_m. \end{aligned}$$

So, if we want to find u_k for some time step k , far into the future, we get

$$\begin{aligned} u_k &= E^k u_0 = \alpha_1 E^k v_1 + \alpha_2 E^k v_2 + \dots + \alpha_m E^k v_m \\ &= \alpha_1 \lambda_1^k v_1 + \alpha_2 \lambda_2^k v_2 + \dots + \alpha_m \lambda_m^k v_m. \end{aligned}$$

With this we can predict long term behavior of the diffusion. (You will get a chance to do this in the exercises later.)

20.1 Eigenvectors and Eigenvalues

This property that the special vectors in a heat diffusion have is a very desirable property elsewhere. So, in Linear Algebra, we give these vectors a name.

Definition 20.1. *Let V be a vector space. Given a linear operator $L : V \rightarrow V$, with corresponding square matrix, and a nonzero vector $v \in V$. If $Lv = \lambda v$ for some scalar λ , then we say v is an eigenvector of L with eigenvalue λ .*

As with the heat states, we see that eigenvectors (with positive eigenvalues) of a linear operator only change amplitude when the operator is applied to the vector. This makes repetitive applications of a linear operator to its eigenvalues very simple.

Notice that in order to find the eigenvalues and eigenvectors of an operator, we need only solve the equation

$$(L - \lambda I)v = 0.$$

Since every linear operator has an associate matrix, we will treat L as a matrix for the rest of this section. We want to find nonzero solutions to this equation. That means we want to find the scalars λ so that

$$\det(L - \lambda I) = 0. \tag{8}$$

We call Equation (8) the characteristic equation and $\det(L - \lambda I)$ the characteristic polynomial of L .

Let us now look at some examples:

Example 20.1. *Let us find the eigenvalues and eigenvectors of the matrix*

$$A = \begin{pmatrix} 5 & -3 \\ 6 & -4 \end{pmatrix}.$$

First, we solve the characteristic equation for λ :

$$\begin{aligned} \det(A - \lambda I) &= 0 \\ \left| \begin{pmatrix} 5 & -3 \\ 6 & -4 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| &= 0 \\ \left| \begin{matrix} 5 - \lambda & -3 \\ 6 & -4 - \lambda \end{matrix} \right| &= 0 \\ (5 - \lambda)(-4 - \lambda) + 18 &= 0 \\ \lambda^2 - \lambda - 2 &= 0 \\ (\lambda + 1)(\lambda - 2) &= 0 \\ \lambda = -1, \lambda = 2 \end{aligned}$$

So, now we have two eigenvalues $\lambda_1 = -1$ and $\lambda_2 = 2$. (In case it isn't clear, the subscripts on the λ 's have nothing to do with the actual eigenvalue. I'm just numbering them.) Using our eigenvalues and Equation 6, we can find the corresponding eigenvectors. Let's start with $\lambda_1 = -1$. We want to find v so that

$$(A - (-1)I)v = 0.$$

We can set this up as a system of equations:

$$\begin{aligned} (A + I)v &= 0 \\ \begin{pmatrix} 6 & -3 \\ 6 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ 6x + -3y &= 0 \\ 6x + -3y &= 0 \end{aligned}$$

We can see that this system has infinitely many solutions (that's what we expected) and the solution space is

$$\mathcal{E}_1 = \left\{ \begin{pmatrix} x \\ 2x \end{pmatrix} \middle| x \in \mathbb{R} \right\} = \text{span} \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}.$$

Using the same process, we can find the eigenvectors corresponding to $\lambda_2 = 2$.

$$\begin{aligned} (A - 2I)v &= 0 \\ \begin{pmatrix} 3 & -3 \\ 6 & -6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ 3x + -3y &= 0 \\ 6x + -6y &= 0 \end{aligned}$$

So, the solutions space to this system is

$$\mathcal{E}_2 = \left\{ \begin{pmatrix} x \\ x \end{pmatrix} \middle| x \in \mathbb{R} \right\} = \text{span} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}.$$

20.2 Eigenbasis

In Heat Equation Lab 2, we saw that it would be nice if we could find a set of “simple vectors” that formed a basis for the space of heat states. That is, we found that we could predict long term behavior if we had an eigenbasis. In this section, we want to explore when it is the case that we have an eigenbasis. Let's first consider what we mean by an eigenbasis. When we are searching for eigenvectors and eigenvalues of an $n \times n$ matrix A , we are really considering the linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ defined by $T(v) = Av$. Then the eigenvectors are vectors in the domain of T that are scaled (by the eigenvalue) when we apply the linear transformation T to them. That is, v is an eigenvector with corresponding eigenvalue $\lambda \in \mathbb{R}$ if $v \in \mathbb{R}^n$ and $T(v) = \lambda v$. An eigenbasis is just a basis for \mathbb{R}^n made up of eigenvectors. We define an eigenbasis more formally here.

Definition 20.2. Given an $n \times n$ matrix A . If A has n linearly independent eigenvectors, v_1, v_2, \dots, v_n then $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ is called an eigenbasis of A for \mathbb{R}^n .

Let's look back at Example 20.1. Notice that if we create a set out of the basis elements for both of the eigenspaces \mathcal{E}_1 and \mathcal{E}_2 , we get the set $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$, which is a basis for \mathbb{R}^2 .

Example 20.2. Consider the matrix $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}$. We want to know if A has an eigenbasis for \mathbb{R}^3 . We begin by finding the eigenvectors and eigenvalues. That is, we want to know for which nonzero vectors v and scalars λ does $Av = \lambda v$. We know that this has a nonzero solution v if $\det(A - \lambda I) = 0$. So, we solve for λ in

$$\begin{vmatrix} 2 - \lambda & 0 & 0 \\ 0 & 3 - \lambda & 1 \\ 0 & 0 & 3 - \lambda \end{vmatrix} = 0.$$

Finding the determinant gives the characteristic polynomial

$$(2 - \lambda)(3 - \lambda)(3 - \lambda) = 0.$$

Thus, for $\lambda_1 = 2$ and $\lambda_2 = 3$, there is a nonzero v so that $Av = \lambda v$. We now find the corresponding eigenvectors (really the corresponding eigenspaces). That is, we want to find the solutions to $(A - 2I)v = 0$ and $(A - 3I)v = 0$.

$\lambda_1 = 2$ In this case we are solving the matrix equation

$$(A - 2I)v = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0.$$

So, we will reduce the corresponding augmented matrix.

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

Thus, a is a free variable and $b = c = 0$. Thus, $\mathcal{E}_1 = \text{span} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$.

$\lambda_2 = 3$ Now, in this case we are solving the matrix equation

$$(A - 3I)v = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0.$$

So, we will reduce the corresponding augmented matrix.

$$\left(\begin{array}{ccc|c} -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right).$$

Thus, b is a free variable and $a = c = 0$. Thus, $\mathcal{E}_2 = \text{span} \left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$.

Notice, that in this example, we only have two linearly independent eigenvectors, $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

and $v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ for in order to get another eigenvector, we would need it to be either in \mathcal{E}_1

(and therefore a scalar multiple of v_1) or in \mathcal{E}_2 (and therefore a scalar multiple of v_2). This means that we cannot find an eigenbasis of A for \mathbb{R}^3 .

Notice that in each example above, the set made by taking the union of each basis for each eigenspace is a linearly independent set. Let's see why this is always true.

Lemma 20.1. *Let A be a matrix and let v be an eigenvector with eigenvalue λ . Then for any scalar c , cv is an eigenvector with eigenvalue λ .*

Proof:(See Exercise 1 below)

Theorem 20.1. *Let A be a matrix and let v_1 and v_2 be eigenvectors with eigenvalues λ_1 and λ_2 respectively. If $\lambda_1 \neq \lambda_2$ then $\{v_1, v_2\}$ is linearly independent.*

Proof: Suppose v_1 and v_2 are nonzero eigenvectors with eigenvalues λ_1 and λ_2 respectively. Suppose also that $\lambda_1 \neq \lambda_2$. Then $Av_1 = \lambda_1 v_1$ and $Av_2 = \lambda_2 v_2$. By Lemma 20.1, αv_1 and βv_2 are eigenvectors with eigenvalues λ_1 and λ_2 respectively. Let's look at two cases 1.) $\lambda_1 = 0$ and $\lambda_2 \neq 0$ (the case when $\lambda_2 = 0$ and $\lambda_1 \neq 0$ is proved similarly so we won't prove it) and 2.) $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$:

Case 1: If $\lambda_1 = 0$ then by definition, $v_1 \in \text{null}A$. But since $\lambda_2 \neq 0$, $v_2 \notin \text{null}(A)$. Suppose that

$$\alpha v_1 + \beta v_2 = 0.$$

Then

$$A(\alpha v_1 + \beta v_2) = 0.$$

But this means that $\beta Av_2 = 0$. So $\beta = 0$. But then $\alpha v_1 = 0$ and so $\alpha = 0$ and $\{v_1, v_2\}$ is linearly independent.

Case 2: If $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$ and

$$\alpha v_1 + \beta v_2 = 0.$$

Then

$$A(\alpha v_1 + \beta v_2) = 0$$

tells us that

$$\alpha \lambda_1 v_1 = -\beta \lambda_2 v_2.$$

Then $\alpha v_1 = \frac{\lambda_2}{\lambda_1} \beta v_2$. Thus, $\frac{\lambda_2}{\lambda_1} \beta v_2$ is an eigenvector with eigenvalue λ_1 and by Lemma 20.1, $\frac{\lambda_2}{\lambda_1} \beta v_2$ is an eigenvector with eigenvalue λ_2 . So

$$A \frac{\lambda_2}{\lambda_1} \beta v_2 = \lambda_2 \beta v_2 \quad \text{and} \quad A \frac{\lambda_2}{\lambda_1} \beta v_2 = \frac{\lambda_2^2}{\lambda_1} \beta v_2.$$

So,

$$\lambda_2 \beta v_2 = \frac{\lambda_2^2}{\lambda_1} \beta v_2 \quad \text{and} \quad (\lambda_1 - \lambda_2) \beta v_2 = 0.$$

Since, $\lambda_2 \neq \lambda_1$ and $v_2 \neq 0$, we see that $\beta = 0$. Thus,

$$A(\alpha v_1 + \beta v_2) = 0$$

implies $\alpha \lambda_1 v_1 = 0$. So, $\alpha = 0$. Therefore $\{v_1, v_2\}$ is linearly independent.

□

20.3 Diagonalizable matrices

We found in the second heat equation lab that if we have a basis made up of eigenvectors, life is good. Well, that's maybe a bit over reaching. What we found was that if $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ is an eigenbasis for \mathbb{R}^n corresponding to the diffusion matrix E then we can write any initial heat state vector $v \in \mathbb{R}^n$ as

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n.$$

Suppose these eigenvectors have eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, respectively. Then with this decomposition into eigenvectors, we can find the heat state at any later time (say k time steps later) by multiplying the initial heat state by E^k . This became an easy computation with the above decomposition because it gives us, using the linearity of matrix multiplication,

$$E^k v = E^k(\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n) = \alpha_1 \lambda_1^k v_1 + \alpha_2 \lambda_2^k v_2 + \dots + \alpha_n \lambda_n^k v_n. \quad (9)$$

We can then apply our knowledge of limits from Calculus here to find the long term behavior. That is, the long term behavior is

$$\lim_{k \rightarrow \infty} \alpha_1 \lambda_1^k v_1 + \alpha_2 \lambda_2^k v_2 + \dots + \alpha_n \lambda_n^k v_n.$$

We see that this limit really depends on the size of the eigenvalues. But we also saw, in Lab 2, that if we change basis to an eigenbasis, the diffusion works out really nice also. Let's remind ourselves how we went about that. First, we notice that if we want all computations in the eigenbasis, we have to reconsider the diffusion matrix transformation as well. That means, we want the matrix transformation that does the same thing that $T(v) = Ev$ does, but this matrix is created using the eigenbasis. That is, we want the matrix representation for the linear transformation that takes a coordinate vector $[v]_{\mathcal{E}}$ (where $\mathcal{E} = \{v_1, v_2, \dots, v_n\}$ is the eigenbasis) and maps it to $[Ev]_{\mathcal{E}}$. Let's call this matrix E' . What we are saying is that we want $E'[v]_{\mathcal{E}} = [Ev]_{\mathcal{E}}$. As always, the columns of this matrix are the vectors that are the result of applying the transformation to the current basis elements (in \mathcal{E}). Thus, the columns of E' are $[Ev_1]_{\mathcal{E}}, [Ev_2]_{\mathcal{E}}, \dots, [Ev_n]_{\mathcal{E}}$. But

$$[Ev_1]_{\mathcal{E}} = [\lambda_1 v_1]_{\mathcal{E}} = \lambda_1 [v_1]_{\mathcal{E}} = \lambda_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$[Ev_2]_{\mathcal{E}} = [\lambda_2 v_2]_{\mathcal{E}} = \lambda_2 [v_2]_{\mathcal{E}} = \lambda_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

$$\begin{aligned} & \vdots \\ [Ev_n]_{\mathcal{E}} &= [\lambda_n v_n]_{\mathcal{E}} = \lambda_n [v_n]_{\mathcal{E}} = \lambda_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \end{aligned}$$

So, we found

$$E' = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \\ \dots & & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}$$

Knowing that a change of basis is a linear transformation (actually, an isomorphism), we can find the matrix representation (usually known as a change of basis matrix). Let's call this matrix M and let's see how this works. We know that $M[v]_{\mathcal{E}} = [v]_{\mathcal{S}}$. This means that if we are given a coordinate vector with respect to the basis \mathcal{E} , this transformation will output a coordinate vector with respect to the standard basis. Recall, to get the coordinate vector in the new basis, we solve for the coefficients in

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n.$$

Then

$$[v]_{\mathcal{E}} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Our favorite way to solve this is to set up the matrix equation

$$\begin{pmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} | \\ v \\ | \end{pmatrix}.$$

Notice that this is the transformation written in matrix form!!!! That is, this means that the matrix representation that takes a coordinate vector with respect to the basis \mathcal{E} to a coordinate vector with respect to the standard basis is

$$\begin{pmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_n \\ | & | & & | \end{pmatrix}.$$

So the matrix representation for the transformation that changes from the eigenbasis \mathcal{E} to the standard basis is given by M . Let's use that to rewrite

$$E'[u]_{\mathcal{E}} = [v]_{\mathcal{E}}.$$

That is, $[u]_{\mathcal{E}} = M^{-1}u$ and $[v]_{\mathcal{E}} = M^{-1}v$ for some u and v in the standard basis. So, we have:

$$M^{-1}(u(t + \Delta t)) = M^{-1}(Eu(t)) = E'M^{-1}u(t)$$

$$u(t + \Delta t) = Eu(t) = ME'M^{-1}u(t)$$

$$u(t + k\Delta t) = E^k u(t) = M(E')^k M^{-1}u(t)$$

$$u(t + \Delta t) = E^k u(t) = M \begin{bmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2^k & & \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n^k \end{bmatrix} M^{-1}u(t)$$

Of course, all of this is dependent on having an eigenbasis for \mathbb{R}^n . Exercise 5 below gives the necessary tools to show that we indeed have an eigenbasis for the diffusion.

Following the same procedure in a general setting, let us see what this means in any context. That is, we want to know when we can actually decompose a matrix A into a matrix product MDM^{-1} where M is invertible (clearly) and D is diagonal. Notice from above we see that to form the columns of M we use the eigenvectors of A . This means that as long as we can

find an eigenbasis $\{v_1, v_2, \dots, v_n\}$, then $M = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_n \\ | & | & & | \end{pmatrix}$ is invertible.

Definition 20.3. Given an $n \times n$ matrix A , we say that A is diagonalizable if there exist matrices M and D so that $A = MDM^{-1}$.

Let's look at an example.

Example 20.3. Let $A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 3 \\ 1 & 0 & 1 \end{pmatrix}$. We want to determine whether A is diagonalizable.

To do this, we need to find the eigenvalues and eigenvectors of A . That is, we want to solve $Av = \lambda v$ for both v and λ . It tends to be easier to find λ first. So, that's what we will do. $(A - \lambda I)v = 0$ has infinitely many solutions when $\det(A - \lambda I) = 0$. In our case,

$$\det(A - \lambda I) = \left| \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 3 \\ 1 & 0 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right| = \left| \begin{pmatrix} 1 - \lambda & 0 & 1 \\ 2 & 1 - \lambda & 3 \\ 1 & 0 & 1 - \lambda \end{pmatrix} \right|$$

$$= (1 - \lambda) \begin{vmatrix} 1 - \lambda & 1 \\ 1 & 1 - \lambda \end{vmatrix} = (1 - \lambda) ((1 - \lambda)^2 - 1) \stackrel{set}{=} 0$$

Thus, $\lambda = 0, 1, 2$. Now, because there are three eigenvalues, Theorem 20.1 tells us that there are at least three eigenvectors that are linearly independent. Thus, A is diagonalizable. Let's find the eigenbasis and decompose A into MDM^{-1} .

$\lambda = 0$: We want to find v so that $(A)v = 0$. We will solve this equation by reducing an augmented matrix.

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 2 & 1 & 3 & 0 \\ 1 & 0 & 1 & 0 \end{array} \right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \Rightarrow v \in \left\{ c \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \mid c \in \mathbb{R} \right\}.$$

$\lambda = 1$: We want to find v so that $(A - I)v = 0$. Again, we will solve this equation by reducing an augmented matrix.

$$\left(\begin{array}{ccc|c} 0 & 0 & 1 & 0 \\ 2 & 0 & 3 & 0 \\ 1 & 0 & 0 & 0 \end{array} \right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \Rightarrow v \in \left\{ c \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \mid c \in \mathbb{R} \right\}.$$

$\lambda = 2$: We want to find v so that $(A - 2I)v = 0$. Once again, we will solve this equation by reducing an augmented matrix.

$$\left(\begin{array}{ccc|c} -1 & 0 & 1 & 0 \\ 2 & -1 & 3 & 0 \\ 1 & 0 & -1 & 0 \end{array} \right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -5 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \Rightarrow v \in \left\{ c \begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix} \mid c \in \mathbb{R} \right\}.$$

From these, we can form the eigenbasis

$$\left\{ \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix} \right\}.$$

We also know

$$M = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 1 & 5 \\ 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

I leave it to the reader to find M^{-1} and show that $MDM^{-1} = A$.

Notice that in Exercise 20.2, there are only two eigenvectors and two eigenvalues. This means that we cannot form an eigenbasis for \mathbb{R}^3 . Thus, A is not diagonalizable.

This might lead someone to think that we can just count the eigenvalues instead of eigenvectors. Let's see an example where this is not the case.

Example 20.4. Let $A = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ 2 & 2 & 2 \end{pmatrix}$. We can find the eigenvalues and eigenvectors of A to determine if A is diagonalizable. Let's step through the same steps. First, we solve the characteristic polynomial $\det(A - \lambda I) = 0$.

$$\begin{vmatrix} 3 - \lambda & 1 & 0 \\ 1 & 3 - \lambda & 0 \\ 2 & 2 & 2 - \lambda \end{vmatrix} = (2 - \lambda) \begin{vmatrix} 3 - \lambda & 1 \\ 1 & 3 - \lambda \end{vmatrix} = (2 - \lambda) ((3 - \lambda)^2 - 1) \stackrel{\text{set}}{=} 0.$$

So, $\lambda = 2, 4$. We only have two eigenvalues! Let's find the corresponding eigenspaces.

$\lambda = 4$: We will solve $(A - 4I)v = 0$.

$$\left(\begin{array}{ccc|c} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 2 & 2 & -2 & 0 \end{array} \right) \rightarrow \left(\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 0 & 4 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & -1/2 & 0 \\ 0 & 1 & -1/2 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right).$$

So, the eigenspace is span $\left\{ \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \right\}$.

$\lambda = 2$: We will solve $(A - 2I)v = 0$.

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

So, the eigenspace is span $\left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$.

Notice that even though we have two eigenvalues, we still have three linearly independent eigenvectors! So, A is diagonalizable with $A = MDM^{-1}$ where

$$M = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Exercises

1. Prove Lemma 20.1.
2. Why does Theorem 20.1 tell us that given a matrix with eigenspaces $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_k$, $\mathcal{E}_1 \cup \mathcal{E}_2 \cup \dots \cup \mathcal{E}_k$ is a linearly independent set?
3. Show that $MDM^{-1} = A$ in Example 20.3.
4. Determine which of the following matrices is diagonalizable. Whenever it is, write out the diagonalization of A .

(a) $A = \begin{pmatrix} 1 & 1 \\ -2 & 3 \end{pmatrix}$

(b) $A = \begin{pmatrix} 5 & 10 & -3 \\ 0 & -2 & 0 \\ 0 & 3 & -1 \end{pmatrix}$

(c) $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 5 & 4 & 2 \end{pmatrix}$

(d) $A = \begin{pmatrix} 3 & -4 & 2 \\ 2 & -3 & 2 \\ 0 & 0 & 1 \end{pmatrix}$

(e) $A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 5 & 4 & 2 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$

5. Consider the heat diffusion operator $E: R^m \rightarrow R^m$ with standard basis matrix representation

$$E = \begin{bmatrix} 1 - 2\delta & \delta & 0 & 0 & & \\ & \delta & 1 - 2\delta & \delta & 0 & \dots \\ & 0 & \delta & 1 - 2\delta & \delta & \\ & & & \vdots & & \ddots \end{bmatrix},$$

where $0 < \delta < \frac{1}{4}$. Show that the k^{th} eigenvector v_k ($1 \leq k \leq m$) is given by

$$v_k = \left(\sin \frac{\pi k}{m+1}, \sin \frac{2\pi k}{m+1}, \sin \frac{3\pi k}{m+1}, \dots, \sin \frac{(m-1)\pi k}{m+1}, \sin \frac{m\pi k}{m+1} \right)$$

and provide the k^{th} eigenvalue. Discuss the relative size of the eigenvalues.

21 Heat Equation for Linear Algebra Lab #3

In this lab, you will explore details of the nature of heat diffusion. This lab requires that you use Matlab or Octave and gives you commands to type at the Matlab/Octave prompt.

1. First watch the demo from class of the diffusion on a heat state by typing the following command in Matlab/Octave:

```
HeatEqnClassDemos(1);
```

What characteristics of heat flow (discussed at the beginning of this module) did you observe in the demo? In the rest of this lab, you'll discover how those characteristics can be traced back to linear algebra concepts that we've been studying!

2. First start by finding the eigenvectors and eigenvalues for the heat diffusion operator. Do this using the `eig` command in MatLab (or Octave). First, run the function that creates the diffusion operator by typing, at the Matlab/Octave prompt, the two lines below (with a "return" or "enter" key).

```
m=5;
E=full(EvolutionMatrix(m));
```

Check that E looks like you want it to by typing

```
E
```

What is the value of δ used by this code? Now, use then use the `eig` command to find the eigenvalues and eigenvectors

```
[V,D]=eig(E);
```

3. Now, we want to verify that V is the matrix whose columns are eigenvectors and D is the matrix whose diagonal entries are the eigenvalues.
 - (a) To see that D is actually diagonal, type

D

- (b) Now verify that the first column of V is the eigenvector of E whose eigenvalue is the first diagonal entry of D . Do this by typing in Matlab/Octave

```
E*V(:,1)
D(1,1)*V(:,1)
```

The first of these commands multiplies our diffusion operator E by the first column of the V matrix. The second of these commands multiplies the $(1, 1)$ entry of the D matrix by the first column of V . These should be the same.

- (c) Try this again with the second column of V and the $(2, 2)$ entry of D . You may notice that the third entry in this eigenvector may be represented by a very small value $\sim 10^{-16}$. This is a numerical artifact; such small values in relation to other entries should be taken to be zero.
- (d) Now we can get a better form for the entries in D . Type

```
L=diag(D)
```

This gives a vector made of the diagonal elements of D . (Caution: the `diag` command has many uses other than extracting diagonal elements.)

4. Let's see what this has to do with our heat diffusion.

- (a) Now redo the steps (2) and (3d) in Matlab/Octave with `m=100` to get the new eigenvectors and eigenvalues of E .
- (b) Now we will visualize these eigenvectors as heat states.
- We will choose to view 5 eigenvectors with their the corresponding eigenvalues. Use the plotting function `EigenStuffPlot` by typing

```
choices=[80,85,90,95,100];
EigenStuffPlot(V(:,choices),L(choices));
```

- How are the individual eigenvectors similar or dissimilar? Make some observations about the relationship between these eigenvectors and eigenvalues.
- Choose different eigenvectors (remember we have 100 of them) to plot by changing entries in vector in the command that starts

```
choices=...
```

What did you choose for `choices`?

- Now rerun the plotting function by entering

```
EigenStuffPlot(V(:,choices),L(choices));
```

- v. Write a list of observations relating eigenvectors and eigenvalues. (Try the above commands with a few more choices if needed to answer this question.)

5. In the previous lab, you wrote an arbitrary heat state as a linear combination of the eigenvectors. Let's view some diffusions of linear combinations of a few eigenvectors.

(a) First, just two eigenvectors.

- i. Choose the eigenvectors:

```
choices=[60,80];
```

- ii. Choose their coefficients (these should be between -2 and 2 for the stability of the code):

```
coeffs=[1,-0.25];
```

- iii. What is the linear combination we are about to plot?
- iv. Plot and watch the diffusion of these by choosing a maximum number of time steps `MaxTime` and running the function `DiffuseLinearCombination`:

```
MaxTime=50;  
DiffuseLinearCombination(V(:,choices),L(choices),coeffs,MaxTime);
```

- v. Do this again by changing the command lines starting as

```
choices=...  
coeffs=...
```

You can change `MaxTime` if you want also, but making it larger than 500 might make you sit for a very long time. You can also change the pause time between frames by giving a fifth input to function `DiffuseLinearCombination` which is the pause time in seconds.

- vi. Make some observations about the diffusion of a linear combinations of two eigenvectors. Try various linear combinations as needed.

(b) Now, 5 eigenvectors (the code is not set up to do more).

- i. Choose the eigenvectors:

```
choices=[60,70,80,90,100];
```

- ii. Choose their coefficients (these should be between -2 and 2 for the stability of the code):

```
coeffs=[1,-1,1,-1,1];
```

- iii. What is the linear combination we are about to plot?
- iv. Plot and watch (you may want to watch this multiple times) the diffusion of these by choosing a maximum number of time steps `MaxTime` and running the function `DiffuseLinearCombination`:

```
MaxTime=100;  
DiffuseLinearCombination(V(:,choices),L(choices),coeffs,MaxTime);
```

- v. Do this again by changing the command lines starting as

```
choices=...  
coeffs=...
```

- vi. Make some observations about the diffusion of linear combinations of eigenvectors. Try various linear combinations as needed.
- (c) Use the above explorations to make a statement about diffusion details for an arbitrary heat state $u = \alpha_1\beta_1 + \alpha_2\beta_2 + \dots + \alpha_m\beta_m$, where the β_i are eigenvectors.

22 Radiography and Tomography in Linear Algebra

Lab #4

In this activity, you will explore some of the properties of radiographic transformations.

In Lab #3 you investigated the questions: (1) whether a radiographic operator can produce the same radiographs from different objects, (2) whether there are invisible objects other than the “zero” object, and (3) whether there are radiographs that are not the image of any object.

22.1 Task 1

After reviewing Lab 3, answer the following questions about spaces related to a radiographic transformation T .

1. For a radiographic transformation T , is the nullspace of T a subspace consisting of objects or of radiographs?
2. Describe the importance of the nullspace of a radiographic transformation given that the goal is to reconstruct objects.
3. For a radiographic transformation T , is the column space of T a subspace consisting of objects or of radiographs?
4. Describe the importance of the column space of a radiographic transformation given that the goal is to reconstruct objects.

22.2 Task 2

Now consider the questions of whether T is injective and/or surjective.

1. What does it mean for T to be injective?
2. What does it mean for T *not* to be injective? If two objects produce the same radiograph, what can you say about the ‘object’ that is their difference?
3. What does it mean for T to be surjective? Not surjective?

22.3 Task 3

Using your answers above, consider the following questions.

1. What must be true about the radiographic transformation T so that can we determine (reconstruct) the object that produces a given radiograph?
2. Suppose that the matrix corresponding to the radiographic transformation T is injective and square. In order to reconstruct an object that produces a given radiograph, what must be true about the transformation?
3. Is it possible for the radiographic transformation to correspond to a nonsquare matrix that is taller than it is wide (more rows than columns)? If so, what would this mean? If not, why not?
4. Is it possible for the radiographic transformation to correspond to a non-square matrix that is wider than it is tall (more columns than rows)? If so, what would this mean? If not, why not?
5. Brainstorm ideas that might help to reconstruct an object for a given radiograph if the radiographic transformation is injective, but the corresponding matrix is not square.

23 Special Matrices

Let’s pause and explore some matrices that we may have seen in this course already. We want to review some properties of these matrices.

23.1 Identity Matrix

The Identity matrix is the square matrix whose entries are zero except along the diagonal, where they are ones. These matrices has been used as a goal when reducing matrices. When we reduce a matrix, the identity is the form we have in mind, whether or not we actually achieve that form is not the issue. We use the identity to test inverses: A is the inverse of B if $AB = I$ and $BA = I$. When finding the inverse of a matrix A , we augment the matrix with the identity ($[A|I]$) and reduce so that the left side of the augment is the identity ($[I|A^{-1}]$). The identity matrix is the easiest (except, of course, the zero matrix) matrix to multiply by.

23.2 Diagonal Matrices

Diagonal matrices are similar to the identity with the only possible difference being along the diagonal. The diagonal matrix can have any entry (even zero). Diagonal matrices are nice to have when we want to repeatedly multiply by one. Indeed, if

$$D = \begin{pmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & d_n \end{pmatrix},$$

then

$$D^k = \begin{pmatrix} d_1^k & 0 & 0 & \dots & 0 \\ 0 & d_2^k & 0 & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & d_n^k \end{pmatrix}.$$

Multiplying by a diagonal matrix is almost as easy as multiplying by the identity. In this case, the components are scaled by the diagonal entries. That is, if

$$v = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix}, \quad \text{then} \quad Dv = \begin{pmatrix} d_1 a_1 \\ d_2 a_2 \\ \dots \\ d_n a_n \end{pmatrix}.$$

23.3 Triangular and Block matrices

There are two types of triangular matrices.

Definition 23.1. Upper triangular matrices are matrices whose entries below the diagonal are all zero. That is, U is an upper triangular matrix if it is of the form

$$U = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \dots & a_{1,n} \\ 0 & a_{2,2} & a_{2,3} & \dots & a_{2,n} \\ 0 & 0 & a_{3,3} & \dots & a_{3,n} \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & a_{n,n} \end{pmatrix}.$$

Lower triangular matrices are matrices whose entries above the diagonal are all zero. That is, L is a lower triangular matrix if it is of the form

$$L = \begin{pmatrix} a_{1,1} & 0 & 0 & \dots & 0 \\ a_{2,1} & a_{2,2} & 0 & \dots & 0 \\ a_{3,1} & a_{3,2} & a_{3,3} & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \dots & a_{n,n} \end{pmatrix}.$$

In both of the above, we see that the matrices are split into two triangles, but named by where the nonzero entries are located. Notice that a diagonal matrix is both upper and lower triangular.

Another type of matrix showing geometric patterns in the layout of the entries are block matrices. Given matrices A_1, A_2, A_3 , and A_4 , we can create the block matrix

$$B = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}$$

as long as A_1 has the same number of rows as A_2 and the same number of columns as A_3 . A_2 also needs to have the same number of columns as A_4 and A_3 must have the same number of rows as A_4 . So, if A_1 is $n_1 \times m_1$ and A_4 is $n_4 \times m_4$, we know that A_2 is $n_1 \times m_4$ and A_3 is $n_4 \times m_1$. We also know that B is $(n_1 + n_4) \times (m_1 + m_4)$. Notice that we can multiply block matrices as if they are regular matrices.

Example 23.1. *Let*

$$B = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} \tilde{A}_1 & \tilde{A}_2 \\ \tilde{A}_3 & \tilde{A}_4 \end{pmatrix}.$$

Then

$$BC = \begin{pmatrix} A_1\tilde{A}_1 + A_2\tilde{A}_3 & A_1\tilde{A}_2 + A_2\tilde{A}_4 \\ A_3\tilde{A}_1 + A_4\tilde{A}_3 & A_3\tilde{A}_2 + A_4\tilde{A}_4 \end{pmatrix}.$$

So as long as the matrix multiplies inside make sense, this matrix multiply also makes sense.

Exercises

- Using the above definition, create a 3×3 upper triangular matrix. Call it U . Now create a 3×3 lower triangular matrix. Call it L . (Hint for later: Don't put too many large numbers because you will use it in upcoming exercises.)
- Find U^2 , U^3 , and U^4 . What do you notice about what happens when you raise U to powers?
- Find L^2 , L^3 , and L^4 . What do you notice about what happens when you raise U to powers?
- Now create a 6×6 block matrix, with a nine 2×2 block matrices. Call it B_1 .
- Create another 6×6 block matrix, with a nine 2×2 block matrices, but this time make only the three blocks along the diagonal nonzero matrices and the others zeros matrices. Call this block matrix B_2 .
- Find B_2^2 and B_2^3 . What do you notice? Write B_2^k in block form.
- Given the matrices A_1, A_2, A_3 , we can create the block matrix:

$$B = \begin{pmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_4 \end{pmatrix},$$

where 0 represents a block of zeros. Find B^k using your observations from above.

23.3.1 The Transpose and Symmetric Matrices

We saw the transpose in Section 2. Finding the transpose of an $n \times m$ matrix gives a new matrix that is $m \times n$. Let's compare the matrix spaces of the transpose of a matrix to the original matrix. We will do this through exercises. In this section, we mention symmetric matrices as well. These matrices also follow a pattern. The diagonal of such matrices acts as a line of reflection (in a sense) for the matrix entries.

Definition 23.2. Symmetric matrices are matrices whose entries correspond across the diagonal so that $a_{i,j} = a_{j,i}$.

An example of a matrix that is symmetric is the matrix

$$\begin{pmatrix} 1 & 3 & 5 \\ 3 & 2 & 6 \\ 5 & 6 & 8 \end{pmatrix}.$$

Note, we mention these matrices in the same section as the transpose because the transpose of a symmetric matrix is itself. More formally, if S is a symmetric matrix, then $S^T = S$. The following theorem is something that is really useful about the transpose of a matrix.

Theorem 23.1. Let V and W be finite-dimensional vector spaces and $T : V \rightarrow W$ a linear transformation. If $\text{rank}(T) = \dim(V)$ then $T^T T$ is invertible. (Here $T^T T$ is understood to be the matrix representation.)

There are two ways you might want to think about this.

- **Rationale #1:** $T : V \rightarrow W$ having $\text{rank}(T) = \dim(V) = n$ (full rank) gives us that $\text{ran}(T)$ is a subspace of W with the same dimension as V . So, we know that if we define the restriction of T^T to $\text{ran}(T)$ to be $\tilde{T}^T : \text{ran}(T) \rightarrow V$, then \tilde{T}^T , having full rank maps onto V . That is, $\text{ran}(\tilde{T}^T) = V$. So, $T^T T : V \rightarrow V$ has only a trivial null space.
- **Rationale #2:** Show that $N(T^T T) = \{0\}$, then because $T^T T$ is square we have that $T^T T$ is invertible. Let $x \in N(T^T T)$. We have

$$0 = x^T 0 = x^T T^T T x = (Tx)^T (Tx) = \|Tx\|^2$$

which implies $Tx = 0$, that is, $N(T^T T) = N(T)$. Now, since $\text{rank}(T) = \dim(V)$, $N(T) = \{0\}$ and $N(T^T T) = \{0\}$.

Exercises

1. For each matrix, find A^T and then the row, space, column space, nullspace, rank, and nullity of both A and A^T .

(a)

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 5 \\ 2 & 3 & 7 \\ 2 & 1 & 4 \end{pmatrix}.$$

(b)

$$A = \begin{pmatrix} 1 & 0 & 2 & 4 \\ 1 & 1 & 5 & 3 \\ 2 & 1 & 7 & 1 \end{pmatrix}.$$

(c)

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 5 \\ 3 & 2 & 8 \end{pmatrix}.$$

2. What do you notice about how the matrix spaces of A and their dimensions compare to the matrix spaces of A^T and their dimensions? List all patterns you saw.
3. For each A and A^T above, find AA^T and $A^T A$. Find the matrix spaces of this new matrix and their dimensions. What do you notice about how these matrix spaces compare to those of A and A^T ?
4. Do you notice anything about the structure of the matrices AA^T and $A^T A$?

24 Radiography and Tomography in Linear Algebra Lab #5: Reconstruction Without an Inverse

In this lab, we will consider the cases when the matrix representation of the radiographic transformation does not have an inverse. (Note: we say that $T : V \rightarrow W$ has an **inverse** $T^{-1} : W \rightarrow V$ if $T^{-1}T = I_V$, the identity mapping from V to V , and $TT^{-1} = I_W$ the identity mapping from W to W . We say that $T : V \rightarrow W$ has a **one-sided inverse**, $P : W \rightarrow V$ if $PT = I_V$.)

24.1 Invertible transformation

In this section, we will consider the following example: We are given a radiograph with 24 pixels that was created by applying some radiographic transformation, T to an object with 16 voxels.

1. Give a scenario for a radiographic transformation T that fits the above example. Don't calculate a T , rather give the following:
 - Size of the object: $___ \times ___$.
 - Number of pixels per view:
 - Number of views:
2. Suppose we know b and we want to find x . This means that we want

$$x = T^{-1}b.$$

- (a) What properties must the *transformation* T have so that T is invertible?
- (b) What properties must the *transformation* T have so that the one-sided inverse of T exists?

- (c) What matrix properties must the *matrix representation* of T have so that it is invertible?
- (d) When $N \leq M$ (as in the example above), what matrix properties must the *matrix representation* of T have so that it has a one-sided inverse?
3. For ease of notation, we typically use the same notation for the matrix and the transformation, that is, we call the matrix representation of T , T . Suppose, $N \leq M$ and a one-sided inverse, P of T exists. This means that $x = Pb$. We know that if T is invertible, we have that $P = T^{-1}$ and $x = Pb$. But, in the example above, we know that T is not invertible. Using the following steps, find the one-sided inverse of T .
- (a) Because $Tx = b$, for any linear operator A , we can write $ATx = Ab$. This is helpful if AT is invertible. Since T is one-to-one, we know that for AT to be invertible, the only vector in $\text{ran}(T)$ that is in $\text{null}(A)$ is the zero vector. What other properties must A have so that AT is invertible?
- (b) Provide a matrix, A so that A has the properties you listed in 3a and so that AT is invertible.
- (c) Solve for x in the matrix equation $ATx = Ab$ using the A you found and provide a representation of the one-sided inverse of P .
4. Putting this all together now, state the necessary and sufficient condition for T to have a one-sided inverse?

24.2 Application to a small example

Consider the following radiographic example.

- Total number of voxels: $N = 16$ ($n = 4$).
 - Total number of pixels: $M = 24$
 - $ScaleFac = 1$
 - Number of views: $a = 6$
 - View angles: $\theta_1 = 0^\circ, \theta_2 = 20^\circ, \theta_3 = 40^\circ, \theta_4 = 60^\circ, \theta_5 = 80^\circ, \theta_6 = 100^\circ$.
1. Use `tomomap.m` to compute T and verify that the one-sided inverse of T must exist. Note that function `tomomap` returns a transformation matrix in sparse format. To use and view as a full matrix array use the command `T=full(T)`; after constructing T .
 2. Compute the one-sided inverse P . Use P to find the object that created the following radiograph vector (You should be able copy and paste this into Octave or Matlab. If you need assistance in using your one-sided inverse, ask and I will help you with syntax):

```
b=[0.00000
32.00000
32.00000
0.00000
```

```
1.97552
30.02448
30.02448
1.97552
2.71552
29.28448
29.28448
2.71552
2.47520
29.52480
29.52480
2.47520
1.17456
30.82544
30.82544
1.17456
1.17456
30.82544
30.82544
1.17456]
```

24.3 Application to Brain Reconstruction

Now, lets reconstruct some brain images from radiographic data. This section will guide you in this process.

1. Collect the necessary provided files and place them in your working Octave/Matlab directory.
 - (a) Data File: `Lab5radiographs.mat`
 - (b) Plotting Script: `ShowSlices.m`
 - (c) Octave/Matlab Code: `tomomap.m`
2. Choose whether to create a new script file (“`.m`” file) for all of your commands or to work at the Octave/Matlab prompt.
3. Load the provided radiographic data with the following command.

```
load Lab5radiographs.mat
```

This line loads an array variable named `B` which is a `12960x362` array. Each column is a radiograph corresponding to one horizontal slice of the human head. Each radiograph has 12960 total pixels spread across many views. The first 181 columns are noiseless radiographs and the last 181 columns are corresponding radiographs with a small amount of noise added.

4. Use the familiar function `tomomap.m` to construct the transformation operator T corresponding to the following scenario (which is the scenario under which the radiographic

data was obtained): $n = 108$, $m = 108$, $ScaleFac = 1$, and 120 view angles: the first at 1° , the last at 179° , and the rest equally spaced in between (hint: use the `linspace` command).

5. Some Octave/Matlab functions do not work with sparse arrays (such as your T). So, simply just make T a full array with this command:

```
T=full(T);
```

6. It is tempting to compute the one-sided inverse P as found in 3c of the first part. However, such a large matrix takes time to compute and much memory for storage. Instead we can use a more efficient solver provided by Octave/Matlab. If we seek a solution to $Lx = b$, for invertible matrix L , we find the unique solution by finding L^{-1} and then multiplying it by b . Octave/Matlab does both operations together in an efficient way (by not actually computing L^{-1}) when we use the command `x=L\b`.

Let's try this with the first radiograph in the matrix B . That is, we will reconstruct the slice of the brain that produced the radiograph that is represented in the 50th column of B . We want to solve the equation you found 3c (recall the matrix you found to be A). Using the A you found in 3c in the following commands

```
b=B(:,50);  
x=(A*T)\(A*b);
```

you will find your first brain slice reconstruction. Now to view your reconstruction, use the following commands

```
figure;  
x=reshape(x,108,108);  
imshow(x,[0,255]);
```

(The reshape command is necessary above because the result x is a $(108 \cdot 108) \times 1$ vector, but we want the object to be 108×108 image.)

7. Notice also that x and b could be matrices, say X and B . In this case, each column of X is the unique solution (reconstruction) for the corresponding column of B . Use these ideas to reconstruct *all* 362 brain slices using a single Octave/Matlab command. Use the variable name X . Make sure that X is an 11664×362 array. Now, the first 181 columns are reconstructions from noiseless data and the last 181 columns are reconstructions from noisy data.
8. Run the script file `ShowSlices.m` which takes the variable X and plots example slice reconstructions. Open `ShowSlices.m` in an editor and observe the line

```
slices=[50 90 130];
```

You can choose to plot any three slices by changing the slice numbers. In the figure, the left column of images are reconstructions from noiseless data, the right column of images are reconstructions from the corresponding noisy data. **IMPORTANT:** Once you have computed X you need only run `ShowSlices.m` to view different slices; running the other commands described above is time consuming and unnecessary.

Congratulations! You have just performed your first brain scan tomography. Using your new tool, answer the following questions.

1. Are the reconstructions from the noiseless radiographs exact discrete representations of the brain? Based on the concepts you have learned about matrix equations, explain why the reconstructions are exact or why not?
2. Determine a way to find the relative amount of noise in the noisy radiographs. It might be useful to know that the noisy radiograph corresponding to the radiograph represented in the 50th column of B is in the $(181 + 50)^{th}$ column of B . (Note: There are many ways you can determine this, be creative.) Is there a lot or a little noise added to the radiographs?
3. Are the reconstructions from the noisy radiographs exact representations of the brain?
4. Compare the degree of “noisiness” in the noisy radiographs to the degree of “noisiness” in the corresponding reconstructions. Draw some conclusion about what this comparison tells us about the practicality of this method for brain scan tomography.

25 Inner Products and Projections

In this section, we will discuss a more generalized dot product, how we can measure the size of a vector using this generalized dot product, and then we will use these to discuss a particular type of linear transformation, a projection.

25.1 Inner Products

We have already seen the dot product for vectors in \mathbb{R}^n . We want to generalize this notion to all vector spaces. We do this by defining the inner product. Below we give the formal definition.

Definition 25.1. *Let V be a vector space with scalars in \mathbb{R} . An inner product is a mapping $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ that satisfies the following three properties. For every $u, v, w \in V$*

1. $\langle u, v \rangle = \langle v, u \rangle$ (Symmetric)
2. $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ (Linearity in the first argument)
3. $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ if and only if $x = 0$. (Positive definite)

A vector space with an inner product is called an inner product space.

It is important to note that the above definition assumes our scalars are real. If they are not, the symmetry property can be different, but for now, we will consider only real inner product spaces. Let's look at an example:

Example 25.1. *Let us define $\langle \cdot, \cdot \rangle : \mathcal{M}_{2 \times 3} \times \mathcal{M}_{2 \times 3} \rightarrow \mathbb{R}$ by $\langle u, v \rangle$ the sum of the component-wise products. That is, if*

$$u = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \quad \text{and} \quad v = \begin{pmatrix} g & h \\ j & k \\ \ell & m \end{pmatrix}$$

then,

$$\langle u, v \rangle = ag + bh + cj + dk + e\ell + fm.$$

Notice that because multiplication in \mathbb{R} is commutative, $\langle u, v \rangle = \langle v, u \rangle$. Also, because multiplication distributes over addition, we get linearity: let

$$w = \begin{pmatrix} n & p \\ q & r \\ s & t \end{pmatrix}.$$

Then

$$\begin{aligned} \langle u + v, w \rangle &= (a + g)n + (b + h)p + (c + j)q + (d + k)r + (e + \ell)s + (f + m)t \\ &= (an + bp + cq + dr + es + ft) + (gn + hp + jq + kr + \ell s + mt) = \langle u, w \rangle + \langle v, w \rangle. \end{aligned}$$

Finally, notice that

$$\langle u, u \rangle = a^2 + b^2 + c^2 + d^2 + e^2 + f^2 \geq 0.$$

And, if $\langle u, u \rangle = 0$ then $a = b = c = d = e = f = 0$ so $u = 0$. So, this is indeed an inner product.

Let's look at another example

Example 25.2. Let $\langle \cdot, \cdot \rangle : \mathcal{P}_2 \times \mathcal{P}_2 \rightarrow \mathbb{R}$ be defined by

$$\langle p_1, p_2 \rangle = \int_0^1 p_1 p_2 \, dx, \quad \text{for } p_1, p_2 \in \mathcal{P}_2.$$

Notice again that because polynomial multiplication is commutative, the symmetric property holds:

$$\langle p_1, p_2 \rangle = \langle p_2, p_1 \rangle.$$

To show linearity in the first argument, we let $p_1 = ax^2 + bx + c$, $p_2 = dx^2 + ex + f$, and $p_3 = gx^2 + hx + k$ and we compute (I've suppressed a lot of algebra here).

$$\begin{aligned} \langle p_1 + p_2, p_3 \rangle &= \int_0^1 (ax^2 + bx + c + dx^2 + ex + f)(gx^2 + hx + k) \, dx \\ &= \int_0^1 ((a + b)gx^4 + ((b + e)g + (a + d)h)x^3 \\ &\quad + ((a + d)k + (c + f)g + (b + e)h)x^2 \\ &\quad + ((b + e)k + (c + f)h)x + (c + f)k) \, dx \\ &= \int_0^1 (agx^4 + (bg + ah)x^3 + (ak + cg + bh)x^2 + (bk + ch)x + ck) \, dx \\ &\quad + \int_0^1 (dgx^4 + (eg + eh)x^3 + (ek + fg + eh)x^2 + (ek + fh)x + fk) \, dx \\ &= \langle p_1, p_3 \rangle + \langle p_2, p_3 \rangle \end{aligned}$$

Finally, if $p \in \mathcal{P}_2$ then $\langle p, p \rangle = \int_0^1 p^2(x) \, dx \geq 0$. If $\langle p, p \rangle = 0$ then $\int_0^1 p^2(x) \, dx = 0$ which means p must be the zero polynomial. So, we now have an inner product on polynomial spaces.

Notice that this was only an example and can be extended to all polynomial spaces. Now that we have inner products, we can define a way to measure the “size” of a vector using the inner product.

Definition 25.2. Let V be an inner product space with inner product $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$. We define the norm of $u \in V$ by $\|u\| = (\langle u, u \rangle)^{1/2}$.

We can compare this to the magnitude of a vector in \mathbb{R}^n . Recall, we said that, when in \mathbb{R}^n , the inner product is the same as the dot product. So, according to Definition 25.2, if $u \in \mathbb{R}^n$ is the vector $u = (u_1, u_2, \dots, u_n)$, then $\|u\| = (\langle u, u \rangle)^{1/2} = (u \cdot u)^{1/2} = \sqrt{u_1^2 + u_2^2 + \dots + u_n^2}$. This is exactly how we have always defined the magnitude of a vector.

Using this definition with Example 25.2, we can find the “size” of a polynomial. Let $p \in \mathcal{P}_2$ be defined by $p = ax^2 + bx + c$. Then

$$\begin{aligned} \|p\| &= \left(\int_0^1 (ax^2 + bx + c)^2 dx \right)^{1/2} \\ &= \left(\int_0^1 a^2x^4 + 2abx^3 + (2ac + b^2)x^2 + 2bcx + c^2 dx \right)^{1/2} \\ &= \left(\frac{a^2}{5} + \frac{ab}{2} + \frac{2ac + b^2}{3} + bc + c^2 \right)^{1/2}. \end{aligned}$$

When we first learn about vectors, in \mathbb{R}^n , we learn that the dot product of two perpendicular vectors is zero. Let’s define this situation more generally.

Definition 25.3. Let V be an inner product space with inner product $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$. Given two vectors $u, v \in V$, we say u and v are orthogonal if and only if $\langle u, v \rangle = 0$.

Notice that this definition tells us that the $0 \in V$ is orthogonal to all other vectors.

Definition 25.4. We say that a set of vectors $\{v_1, v_2, \dots, v_n\}$ is orthogonal if and only if $\langle v_i, v_j \rangle = 0$ whenever $i \neq j$. If an orthogonal set also satisfies $\langle v_i, v_i \rangle = 1$, then we say that the set is orthonormal.

The first part of Definition 25.4, says that the vectors are pairwise orthogonal. The second part of Definition 25.4 says the vectors all have norm (or magnitude) equal to 1. (We say that a vector is normalized or normal if it has norm 1.)

Orthogonality will play a big role in helping us reconstruct objects from radiographs. Let’s first look at some properties of orthogonality.

Theorem 25.1. Let V be an n -dimensional vector space and let $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ be a set of vectors in V . If \mathcal{B} is orthogonal, then \mathcal{B} is a basis for V .

proof: Notice that we need only prove that \mathcal{B} is linearly independent since there are n elements in \mathcal{B} . I leave this proof as an exercise for the reader.

In this class, we will focus our attention specifically on the inner product space, \mathbb{R}^n with the inner product defined to be the dot product. The following will generalize to other inner product spaces and it would be a great use of your time between semesters to see how generalizations can be made rigorous.

25.2 Projections

As always, let V be a vector space. Given a proper subspace $W \subsetneq V$ and a vector $v \in V$ (but not in W), we are interested in finding a vector $w \in W$ that is closest to v . Let's first start with an example in \mathbb{R}^2 to understand the idea.

Example 25.3. Suppose we have a vector $u \in \mathbb{R}^3$ and we want to project u onto another vector $v \in \{(x, y, 0) | x, y \in \mathbb{R}\}$ (the x - y plane). In reality, we are projecting onto the line that is parallel to v : $L = \{\alpha v : \alpha \in \mathbb{R}\}$. That is, we want to apply the transformation $\pi_L : \mathbb{R}^3 \rightarrow L$. Then we are looking for the vector $\text{proj}_v(u)$ shown in the figure on the left in Figure ???. Notationally, we write $\pi_L(u) = \text{proj}_v(u)$.

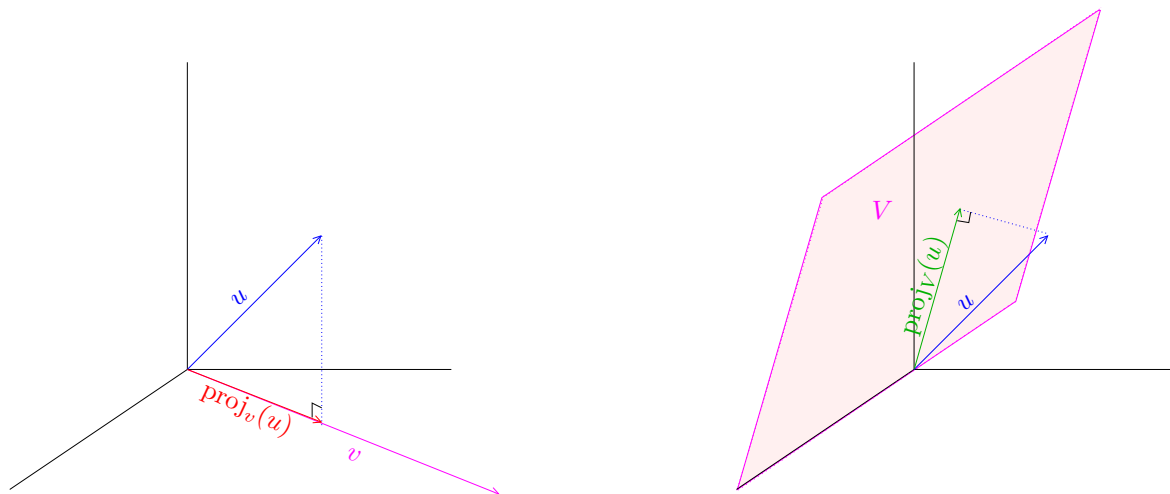


Figure 16: Projection onto a vector (left) and Projection onto a space (right)

Now, suppose we want to project u onto the vector space spanned by two vectors, say $V = \text{span}\{(1, 1, 1), (1, 0, 0)\}$. Then, we want to apply $\pi_V : \mathbb{R}^3 \rightarrow V$ to get $\text{proj}_V(u)$ shown in the figure on the right in Figure 16.

In terms of the Linear Algebra language that we have been using, to find the projection of a vector u onto a space $V = \text{span}\{v_1, v_2, \dots, v_n\}$, we first find the part (or component) of u that is orthogonal to the basis elements of V (thus, orthogonal to all vectors in V). We call this component u_\perp . Then, the projection, $\text{proj}_V(u)$ is the vector that is left over after we subtract the orthogonal part:

$$\text{proj}_V(u) = u - u_\perp.$$

We define this more rigorously below. But first we define some preliminary sets.

Definition 25.5. Let V be a vector space and $W \subset V$. We define the orthogonal complement of W is

$$W^\perp = \{v \in V | \langle v, w \rangle = 0, \text{ for all } w \in W\}.$$

In words W^\perp is just the set of vectors orthogonal to all vectors in W .

Notice that, in Definition 25.5, W is allowed to be the whole space V . This should make you wonder what V^\perp is. In case you are curious, I've made it an exercise.

We use the orthogonal complement to define projections.

Definition 25.6. Let V be a vector space and let W be a subspace of V . If $u \in V$, then we say that the projection of u onto W is the vector in W closest to u . That is,

$$\text{proj}_W(u) = u - n \in W,$$

where n is the vector in W^\perp with smallest norm, $\|n\| = (\langle n, n \rangle)$.

Notice that Definition 25.6 along with Figure 16, we see that, at least in \mathbb{R} , a triangle is formed. Let us use this triangle to compute the projection, $\text{proj}_W(u)$.

We begin with $W = \text{span} \{w_1\}$, the span of only one vector. This means, we want the projection of u onto a line. First, we recognize that we are looking for $w \in W$ so that $u = w + n$ where $n \in W^\perp$. Let's use our knowledge of Linear Algebra to make some conclusions.

- First, $w \in W$ means $w = \alpha w_1$ for some scalar α .
- Next, $n \in W^\perp$ says that $\langle n, w_1 \rangle = 0$.
- Putting these together with $u = w + n$, we get

$$\begin{aligned} u &= \alpha w_1 + n \\ \langle u, w_1 \rangle &= \alpha \langle w_1, w_1 \rangle + 0 \\ \text{So, } \alpha &= \frac{\langle u, w_1 \rangle}{\langle w_1, w_1 \rangle}. \\ \text{Thus, } w &= \frac{\langle u, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1. \end{aligned}$$

The above shows us that as long as we are projecting onto a line, we can easily find the projection. Let's now consider a projection onto a higher dimensional space (such as is seen in Figure 16 on the right). We will follow similar steps we followed in the last example.

This time, we begin with $W = \text{span} \{w_1, w_2, \dots, w_m\}$, where $\{w_1, w_2, \dots, w_m\}$ is linear independent. That means w is in an m -dimensional space. This means, we want the projection of u this m -dimensional space. First, we recognize that we are looking for $w \in W$ so that $u = w + n$ where $n \in W^\perp$. We will attempt to follow the same steps as above.

- First, $w \in W$ means $w = \alpha_1 w_1 + \alpha_2 w_2 + \dots + \alpha_m w_m$ for some scalars $\alpha_1, \alpha_2, \dots, \alpha_m$.
- Next, $n \in W^\perp$ says that $\langle n, w_i \rangle = 0$ for $i = 1, 2, \dots, m$.
- Putting these together with $u = w + n$, we get

$$\begin{aligned} u &= \alpha_1 w_1 + \alpha_2 w_2 + \dots + \alpha_m w_m + n \\ \langle u, w_1 \rangle &= \alpha_1 \langle w_1, w_1 \rangle + \alpha_2 \langle w_2, w_1 \rangle + \dots + \alpha_m \langle w_m, w_1 \rangle + 0 \\ \langle u, w_2 \rangle &= \alpha_1 \langle w_1, w_2 \rangle + \alpha_2 \langle w_2, w_2 \rangle + \dots + \alpha_m \langle w_m, w_2 \rangle + 0 \\ &\vdots \\ \langle u, w_m \rangle &= \alpha_1 \langle w_1, w_m \rangle + \alpha_2 \langle w_2, w_m \rangle + \dots + \alpha_m \langle w_m, w_m \rangle + 0 \end{aligned}$$

So, to get $\alpha_1, \alpha_2, \dots, \alpha_m$, we need to solve a system of equations with m unknowns and n equations. This looks pretty tedious! It would be so nice if $\langle w_i, w_j \rangle = 0$ whenever $i \neq j$. If this were the case, we could easily solve for $\alpha_1, \alpha_2, \dots, \alpha_m$. But, this means we want the basis for W to be orthogonal.

Exercises

1. Prove Theorem 25.1.

2. Given a vector space V and a subspace $W = \left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} \mid a + b + c = 0 \right\}$, what is W^\perp ?

3. Prove that W^\perp is a vector space.

4. Given an ambient vector space V what is $V^\perp \subset V$?

5. Find a basis to $V = \left\{ \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \mid a + 2c = 0 \right\}$. Use the basis you found to find an orthonormal basis.

6. Given the matrix $A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -2 & 0 \\ 1 & 0 & -1 \end{pmatrix}$. Determine whether or not A is an orthogonal matrix. If not, find a matrix whose column space is the same as the column space of A , but is orthogonal.

7. Given the matrices $A_\theta = \begin{pmatrix} \sin\theta & \cos\theta \\ \cos\theta & -\sin\theta \end{pmatrix}$

(a) Draw the vector $u = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and its images under the transformation $T(u) = Au$ for $\theta = 45^\circ, 90^\circ, 30^\circ, 240^\circ$.

(b) Use your results to state what multiplication by A is doing geometrically.

(c) Show that A is an orthogonal matrix for every value of θ .