Linear Algebra in Euclidean Space

MAT223: Linear Algebra I
University of Toronto, Spring 2025

Last Updated: April 11, 2025

Contents

Preface

Linear algebra is the study of linear equations, vectors, vector spaces (which generalize "flat spaces" like lines and planes), and linear transformations (which are functions between vector spaces which preserve lines). The beauty of linear algebra lies in the interplay between these objects: translating between algebraic and geometric perspectives allows us to leverage the strengths of one to deepen our understanding of another.

This course will focus on the fundamentals of linear algebra in Euclidean space. Focusing our study on Euclidean space will allows us to develop the theory naturally, using geometric intuition in familiar two- and three-dimensional Euclidean space as a foundation. From there, we will imagine a geometry in higher dimensional Euclidean space, exploring how to generalize definitions and results to these new domains while preserving their essential properties. In the next course, MAT224, you will take this abstraction further, removing the notion of vectors from Euclidean spaces entirely.

Abstraction is a hallmark of mathematics. We begin with concepts rooted in our intuitive understanding of the world, and then extract core properties to explore how these concept behaves in a more general setting. By building a theory of linear algebra in the most general setting, we create a robust tool for application across many disciplines. While we will not focus this course on direct applications of this material, we will periodically provide resources in your reading assignments for you to explore how the theory we develop can be applied across different disciplines.

How to use these notes. These lecture notes have been shaped over multiple iterations of this course, drawing from numerous sources (listed below). As a living document, they will continue to evolve throughout the semester. Chapters and sections with an asterisk ('$*$') in the title are currently under development. Once an asterisk has been removed, the section will not be edited. In the (hopefully) rare occasion that a section needs to be edited after the asterisk has been removed, an announcement of the changes will be made on Quercus.

Additional resources. The following list of resources may be used to supplement these lecture notes. It is highly recommended to explore alternative presentations of the material. While these notes will provide a thorough account of what's needed for the course, everyone processes information differently. Engaging with supplemental resources, as well as participating in class activities, will help develop your own personal understanding of the content.

FREE RESOURCES:

(1) **3Blue1Brown**, Linear Algebra. (builds geometric intuition with helpful visualizations)
(2) **Siefken**, Linear Algebra: MAT223 Workbook. Available on Quercus. (down-to-earth presentations, concrete examples)
(3) **Alayont and Schlicker**, Linear Algebra and Applications: An Inquiry-Based Approach. (good for extra practice problems)
(4) **Margalit and Rabinoff**, Interactive Linear Algebra (consise presentation of material, interactive examples)
(5) **Johnston**, Introduction to linear and matrix algebra. (logical presentation of material, clear writing)
(6) **Strang**, Linear Algebra Video Lectures (good resource for those who respond well to lecture; associated textbook listed in "Paid Resources" section below)
(7) **Axler**, Linear Algebra Done Right (abstract presentation, often used as textbook for MAT224)

PAID RESOURCES (suggest you find used copies, or search for pdfs):

(1) Poole's *Linear Algebra: A Modern Introduction* (concise account of material with clear writing), see [**2**]
(2) Strang's *Introduction to Linear Algebra* (builds helpful intuition, accompanies lectures posted above), see [**3**]

List of Notation. The following list can be used as a quick reference for notation that appears in these notes.

$\mathbb{R}$	the set of real numbers
$\mathbb{C}$	the set of complex numbers
$\mathbb{Q}$	the set of rational numbers
$\mathbb{Z}$	the set of integers
$\mathbb{R}^n$	n-dimensional Euclidean space

$\in$	is an element of
$\forall$	for all
$\exists$	there exists
$[\vec{x}]_{\mathcal{B}}$	the coordinates of a vector $\vec{x}$ with respect to the basis $\mathcal{B}$
$\dim(V)$	the dimension of a vector space V
T_A	the linear transformation corresponding to the matrix A, given by $T_A(\vec{x}) = A\vec{x}$.
$\ker F$	the kernel of a function F
$\operatorname{im} F$	the image of a function F
$\mathrm{Nul}(A)$	the null space of a matrix A
$\mathrm{Col}(A)$	the column space of a matrix A
$\mathrm{Row}(A)$	the row space of a matrix A
$A^\top$	the matrix transpose
$\det(A)$	the determinant of A
E_λ	the λ-Eigenspace
χ_A	the characteristic polynomial of A
$M_{\mathcal{C}\leftarrow\mathcal{B}}$	change of basis matrix
$\vec{x}\cdot\vec{y}$	the dot product
$\Vert\vec{x}\Vert$	the norm
$d(\vec{x},\vec{y})$	the distance between vectors $\vec{x}$ and $\vec{y}$
$V^\perp$	the orthogonal complement of a vector space V
$\vec{x}_V$	the orthogonal projection of a vector $\vec{x}$ onto a vector space V
$\mathrm{proj}_{\vec{v}}\vec{u}$	the orthogonal projection of $\vec{u}$ onto the vector space $V = \mathrm{Span}(\vec{v})$.

Chapter 1

Systems of Linear Equations

1.1. Introduction to Systems of Linear Equations

Let's look at some examples to motivate our first topic of the course.

Activity 1.1. The following examples have been adapted from Chapter 2 of [**2**], where interested students can find further applications.

Allocation of Resources. A biologist has three strains of bacteria (I, II, and III), which will be placed in a test tube to feed on three different food sources (A, B, and C). The biologist places 230 grams of Food A, 80 grams of Food B, and 150 grams of Food C into the test tube. Given that each bacterium consumes the amount of food indicated in the table below per day, how many bacteria of each strain can coexist in the test tube and consume all of the food?

	Strain I	Strain II	Strain III
Food A	2 g	2 g	4 g
Food B	1 g	2 g	0 g
Food C	1 g	3 g	1 g

Input-Output Analysis. Three neighbors, each with a vegetable garden, agree to share their produce. One will grow beans, one will grow lettuce, and one will grow tomatoes. The following table shows what percentage of each crop each of the neighbors will receive.

	bean grower	lettuce grower	tomato grower
% beans shared	25%	50%	25%
% lettuce shared	40%	30%	30%
% tomato shared	10%	60%	30%

Given this arrangement, explain how the neighbors are valuing each item of produce so that they all break even in this exchange.

The examples above boiled down to us solving a *system of linear equations*. We have the following definitions.

Definition 1.1. A LINEAR EQUATION in variables $x_1, x_2, \ldots, x_n$ is an equation of the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$, where $a_1, \ldots, a_n$ and b are real numbers. In this class, we will often work with systems of equations in four or less variables. For notational convenience, we'll often use the letters x, y, z, w and to indicate our variables, rather than x_1, x_2, x_3, x_4.

Example 1.2. Observe that equations (1) and (3) below are linear, while (2) and (4) are not.

(1) $2x + y + z = 3$

(2) $x^2 - y = 3$

(3) $\pi^2(x + y) = \dfrac{z}{2}$

(4) $e^x + \sqrt{y} = z^3$

Definition 1.3. A SYSTEM OF LINEAR EQUATIONS is a collection of one or more linear equations in the same variables. A tuple $(s_1, \ldots, s_n) \in \mathbb{R}^n$ is a SOLUTION to a system of linear equations if $(s_1, \ldots, s_n)$ is a solution to every linear equation in the system.

Example 1.4. Observe that $(1, 0, 2, -1)$ is a solution to the system of linear equations

$$\begin{cases} x + y - z - w = 0 \\ x - y + 2z + w = 4 \end{cases}$$

Our goal in this chapter is to come up with an algorithm that can solve any system of of linear equations. In Activity 1.1 you may have been able to solve the systems using "ad-hoc" methods. This is fine for small examples, but is not feasible as a general approach. For example, the input-output analysis discussed in Activity 1.1 could be applied on a much larger scale, where there is an exchange of goods among hundreds of citizens. Modeling this problem would involve a system of hundreds of linear equations in hundreds of variables, which is not manageable by hand. In the following section, we introduce some notation which will help us develop a systematic method to handle large systems of this type.

1.2. The Matrix Representation of a Linear System

Consider the system of linear equations

$$\begin{cases} x - y - z = 1 \\ 2x - 3y - z = 3 \\ -x + y - z = -3 \end{cases}$$

Since our variables must all be the same in any system of linear equations, it's a bit redundant to write them every time. Instead, we could just write down the important pieces of this system into an array

$$\begin{array}{rrrr} 1 & -1 & -1 & 1 \\ 2 & -3 & -1 & 3 \\ -1 & 1 & -1 & -3 \end{array}$$

Observe that the first three columns of this array correspond to the coefficients of our independent variables x, y and z for each of our three equations, and the last column corresponds to the constant on the right-hand side of our linear equations. This array is called a *matrix*, and can be used as a bookkeeping device to help us solve systems of linear equations. Let's give some formal definitions.

Definition 1.5. A MATRIX is any rectangular array of quantities or expressions. The quantities or expressions in a matrix are called its ENTRIES. If a matrix has n rows and m columns, then we call this an $n \times m$ matrix.

In this class, our matrices will typically contain real number entries (or sometimes variables working as placeholders for real number entries). Matrices are often written using soft brackets, such as in the 2×2 matrix below

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

or by using hard brackets, such as in the 3×2 matrix below

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

We'll be using soft brackets throughout these notes, but either one is perfectly fine.

Definition 1.6. Consider a general system of linear equations

$$\begin{cases} a_{11}x_1 + a_{12}x_x + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m. \end{cases}$$

The MATRIX OF COEFFICIENTS corresponding to this system is the $m \times n$ matrix given by

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

The AUGMENTED MATRIX of this system is the $m \times (n+1)$ matrix given by

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right).$$

Note that the augmented matrix accounts for the constants on the right-hand side of our equations, while the coefficient matrix does not. Also note that sometimes people do not draw the bar pictured above in an augmented matrix. The vertical line is just an extra piece of decoration for the sake of clarity, it does not change anything about the entries or size of the matrix.

Example 1.7. Consider again the system of linear equations

$$\begin{cases} x - y - z = 1 \\ 2x - 3y - z = 3 \\ -x + y - z = -3 \end{cases}$$

The matrix of coefficients of this system is the 3×3 matrix

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

and the augmented matrix of this system is the 3×4 matrix

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

Note that we could also write the augmented matrix of this system as

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

As mentioned above, the type of brackets used to enclose the matrix and the decision to include the vertical line or not are just choices for how we decorate our matrix. Two matrices with the same entries are equal, no matter how you decorate them.

Remark 1.8. Note that every matrix has an associated system of linear equations. For example, consider the 3×3 augmented matrix

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

Then, A is the augmented matrix of a system of three linear equations in **two** variables

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

Recall that the last column of an augmented matrix corresponds to the constant coefficients of the corresponding system.

This correspondence between systems of linear equations and matrices is an important idea we'll revisit many times throughout the course.

1.3. Equivalent Systems and Elementary Row Operations

In order to generate an algorithm to solve any system of linear equations, it will be convenient to restrict ourselves to just a few types of operations that we can use to manipulate the system. Note that if we perform any of the following operations, we'll obtain a new system of linear equations, but with the same solution set:

(E1) Interchange two equations of the system;

(E2) Replace an equation by a nonzero multiple of itself;

(E3) Replace one equation by the sum of that equation and a scalar multiple of another equation.

We introduce some terminology to streamline our conversation.

Definition 1.9. Two systems of linear equations are called EQUIVALENT if they have the same solution sets.

Example 1.10. Consider the system of linear equations

$$\begin{cases} x - y - z = 1 \\ 2x - 3y - z = 3 \\ -x + y - z = -3 \end{cases} \tag{1.1}$$

We could apply (E3) to replace the third equation with the sum of the first and third equations to obtain the new system of linear equations

$$\begin{cases} x - y - z = 1 \\ 2x - 3y - z = 3 \\ -2z = -2 \end{cases} \tag{1.2}$$

Observe that Systems (1.1) and (1.2) are equivalent (that is, they have the same solution sets), but System (1.2) is simpler to solve since we've isolated one of our variables.

Let's look at what this elementary operation would look like using our new book keeping device. System (1.1) has augmented matrix

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

Note that replacing the third equation with the sum of the first and third equations is identical to replacing the third *row* of the augmented matrix above with the sum of the first and third *row*, which yields

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

Note that this is precisely the augmented matrix for System (1.2).

More generally, our three elementary operations (E1), (E2), and (E3) can be viewed as operations on the rows of the corresponding augmented matrix. We define the following.

Definition 1.11. The ELEMENTARY ROW OPERATIONS are defined as follows:

(ER1) Interchange two rows;

(ER2) Replace a row by a nonzero scalar multiple of itself.

(ER3) Replace one row by the sum of that row and a scalar multiple of another row;

We say that two matrices A and B are ROW EQUIVALENT if one can be obtained from the other by a sequence of elementary row operations. In this case, we write $A \sim B$.

Noting again that applying an elementary row operation is the same as applying an operation to a system of linear equations, which yields an equivalent system of linear equations, we obtain the following key result:

Theorem 1.12. *If two matrices are row equivalent, then the system of linear equations they represent are equivalent (that is, have the same solution sets).*

In short order, we'll see that these three elementary operations are enough to solve *any* system of linear equations. To do this, let's first think about which systems of linear equations are "simplest" to solve.

1.4. Echelon Forms of a Matrix

Activity 1.2. Find all solutions to the the system of linear equations that have the following augmented matrices

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

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

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

In Activity 1.2, we were either able to directly detect that they system had no solutions, or we were able to solve the system using back substitution. In the latter case, we were able to carry this method through because each of the equations in our system had at least one less variable than the equation above it. This feature gave the corresponding augmented matrices an "inverted staircase" pattern. Let's give some definitions to make sense of this pattern more carefully.

Definition 1.13. The PIVOT (also called the LEADING ENTRY) of a row in a matrix is the leftmost nonzero entry.

Example 1.14. The matrix

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

has pivot 1 in the first row, 2 in the second row, and 3 in the third row.

Definition 1.15. A matrix is said to be in ROW ECHELON FORM if

(1) all rows consisting only of zeros are at the bottom, and
(2) the pivot of each row in the matrix is in a column to the right of the pivot of the row above it.

Definition 1.16. A matrix is said to be in REDUCED ROW ECHELON FORM if the matrix is in echelon form and

(1) the pivot in each nonzero row is 1, and
(2) each pivot is the only nonzero entry in its column.

Activity 1.3. Determine which of the following matrices are in row echelon form, reduced row echelon form, or neither.

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 1 \\ 3 & 4 & -1/2 \end{pmatrix}, \; B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; C = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{pmatrix},$$

$$D = \begin{pmatrix} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; E = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; F = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

Activity 1.4. Give examples of 3×4 matrices that satisfy the following conditions. You should provide a new matrix for each part.

(1) In row echelon form, but not reduced row echelon form.
(2) In reduced row echelon form.
(3) In reduced row echelon form and has exactly one row of zeros.
(4) In reduced row echelon form, has exactly one pivot, and has at least one entry other than 1 or 0.

Example 1.2 demonstrate the following key observation.

Observation 1.17. Any system of linear equations whose matrix is in (reduced) row echelon form can be solved using back substitution.

So, one way to solve any system of linear equations is to find an equivalent system whose augmented matrix is in (reduced) row echelon form. In the next section, we show that this can always be done.

1.5. Gauss-Jordan Elimination

We have the following key result.

Theorem 1.18 (Gauss-Jordan)**.** *Every matrix A is row equivalent to a matrix in row echelon form X. Furthermore, if X is in reduced row echelon form, then this matrix is unique.*

Proof. We'll prove the existence of such a row equivalent matrix using "Gauss-Jordan elimination". This method is attributed to the two German mathematicians Carl Friedrich Gauss and Wilhelm Jordan due to their work in the 1800s, but was previously documented in a Chinese text dating back to around 150 BC. The uniqueness step of this argument is quite a bit more involved, and is believable by example, so we omit the proof of this claim here. Suppose that A is an $n \times m$ matrix.

Step 1: Move all rows of zeros to the bottom. Suppose that A has a row of zeros in Row i where $i \neq n$. Letting j be the largest index so that Row j is not a row of zeros, interchange Rows i and j. Repeating this process for all rows gives a row equivalent matrix A_1 so that all rows consisting entirely of zeros are at the bottom of A_1.

Step 2: Move pivots. For all $i < j$ so that the pivot in R_j is to the left of the pivot in R_i in A_1, interchange R_i with R_j. This gives a row equivalent matrix A_2 so that all pivots are either to the right or below of the pivots in columns above it. Furthermore, A_2 still has all rows of zeros at the bottom of the matrix.

Next, for all i, j so that the pivot in R_i is in the same column as the pivot in R_j in A_2, perform the row operation $R_j - a_j/b_i R_i$, where a_j is the pivot in R_j and b_i is the pivot in R_i. This gives a row equivalent matrix A_3 in row echelon form.

Step 3: Make all pivots equal to 1. For all rows R_i containing a pivot a_i in A_3, replace R_i with $(1/a_i)R_i$. This gives a row equivalent matrix A_4 that is still in row echelon form, and where every pivot is equal to 1.

Step 4: Remove nonzero nonpivot entries from pivot columns. Finally, letting a_i be the pivot in row i, if there exists a nonzero entry b_i in row j and column i (the same column as a_i) replace R_j with $R_j - \frac{a_i}{b_i} R_i$. If necessary, repeat previous steps until the matrix is in reduced row echelon form. □

While the algorithm above is complete, it's often easier to allow yourself some flexibility when row reducing a matrix, as demonstrated in the example below. We'll use the following shorthand:

> $R_i \leftrightarrow R_j$: interchange the ith and jth rows;
>
> cR_i: replace row i by row i times the nonzero constant c; Replace a row by a nonzero scalar multiple of itself.
>
> $R_i + cR_j$: replace row i with row i plus the constant c times row j (note that we've written the row which is being replaced first in the operation).

Example 1.19. Below we demonstrate one method to finding a matrix in reduced row echelon form which is row equivalent to the matrix below

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

We have

$$\begin{aligned}
\begin{pmatrix} 1 & 1 & 2 & 1 \\ 1 & 3 & 6 & 1 \\ 1 & 2 & 4 & 1 \end{pmatrix} &\sim \begin{pmatrix} 1 & 1 & 2 & 1 \\ 1 & 3 & 6 & 1 \\ 0 & 1 & 2 & 0 \end{pmatrix}, \ R_3 - R_1 \\
&\sim \begin{pmatrix} 1 & 1 & 2 & 1 \\ 0 & 2 & 4 & 0 \\ 0 & 1 & 2 & 0 \end{pmatrix}, \ R_2 - R_1 \\
&\sim \begin{pmatrix} 1 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 \end{pmatrix}, \ R_2 - 2R_3 \\
&\sim \begin{pmatrix} 1 & 1 & 2 & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ R_2 \leftrightarrow R_3 \\
&\sim \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ R_1 - R_2
\end{aligned}$$

The uniqueness of the matrix X in the previous theorem allows us to define the following.

Definition 1.20. If A is a matrix with reduced row echelon form equal to X, we call X THE REDUCED ROW ECHELON FORM OF A, and write $\mathrm{rref}(A) = X$.

Remark 1.21. Observe that we now have an algorithm to *find all solutions to any system of linear equations.* We summarize the method below.

(1) Given a system of linear equations, write down its corresponding augmented matrix A.
(2) Use Gauss-Jordan elimination to find its reduced row echelon form X. We will often refer to this step as "row reducing A".
(3) Using back substitution, find all solutions to the system of linear equations represented by X. Since X and A are row equivalent, the system of linear equations they represent have the same solution sets.

We saw an example of how to handle step (3) in Example 1.2. We give a few more definitions to algorithmatize this step further.

Definition 1.22. Let A be a matrix with reduced row echelon form equal to X.

(1) We say the ith column of A is a PIVOT COLUMN if X has a pivot in column i.
(2) We say that x_i is a BASIC VARIABLE if the system of linear equations represented by A if the ith column of A is a pivot column.

(3) We say that x_i is a FREE VARIABLE of the system of linear equations represented by A if the ith column of A is not a pivot column.

Observe that, if our system of linear equations has a free variable, then that system has infinitely many solutions. In this case we can PARAMETERIZE the solution set using our free variables. We demonstrate this method by example below.

Example 1.23. Recall that the matrix

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

has reduced row echelon form

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

Note that the pivot columns of A are columns 1 and 2. Since A represents a system of linear equations in three variables x, y, z, we get that x and y are basic variables, while z is free. Note that, even though the last column of A is a pivot column, this column does not correspond to a free variable, since the last column of an augmented matrix always corresponds to the constant coefficients.

Let's look at what this tells us about our solution set. The system of linear equations that X represents is given by

$$\begin{cases} x = 1 \\ y + 2z = 0 \end{cases}$$

Observe that if we let z (our free variable) be *any real number* then our solution set can be described PARAMETRICALLY as $\{(1, -2z, z) : z \in \mathbb{R}\}$. Note that the existence of a free variable guarantees our solution set has infinitely many solutions, and those solutions may be parameterized by the free variables.

Remark 1.24. It is also completely valid for us to instead choose y to vary over the real numbers instead, and have z be controlled by y. In this case, our solution set could be described parametrically as $\{(1, y, -y/2) : y \in \mathbb{R}\}$. You're welcome to describe your solution sets in whatever way makes the most sense to you, but we'll see throughout the course that it will be convenient to understand our variables that are not fixed as the variables corresponding to the non-pivot columns of our augmented matrix, as demonstrated in the example above.

1.6. The Number of Solutions to Systems of Linear Equations

Our work in the previous sections not only tells us how to describe the solution set to any system, but also gives a quick method to detect the *number* of solutions to any given system. We first define the following.

Definition 1.25. A system of linear equations which has *at least one* solution is called CONSISTENT. A system of linear equations which has *no solutions* is called INCONSISTENT.

The result below is often attributed to the mathematicians Eugéne Rouché and Alfredo Capelli. As this will end up being one of our most referenced results this semester, we'll refer to it as the ROCHÉ-CAPELLI THEOREM.

Theorem 1.26 (Rouché-Capelli)**.** *Suppose that a system of linear equations has augmented matrix A and coefficient matrix C. Then,*

(1) The system is inconsistent if and only if the last column of $\operatorname{rref}(A)$ *has a pivot.*

(2) The system has exactly one solution if and only if the last column of $\operatorname{rref}(A)$ *does not have a pivot, and every column of* $\operatorname{rref}(C)$ *has a pivot.*

(3) The system has infinitely many solutions if and only if the last column of $\operatorname{rref}(A)$ *does not have a pivot and* $\operatorname{rref}(C)$ *has a column without a pivot.*

Proof. Suppose that we have a system of m linear equations in n variables. Then, A is an $m \times (n+1)$ matrix, and C is an $m \times n$ matrix.

For part (1), suppose that $\operatorname{rref}(A)$ has a pivot in the last column. Then, $\operatorname{rref}(A)$ contains a row of the form

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

which represents the linear equation $0 = 1$. Since this equation can never have a solution, the system of linear equations represented by A has no solutions, as needed. Now, since $\operatorname{rref}(C)$ either contains a pivot in every column or not, the case where $\operatorname{rref}(A)$ does not have a pivot in every column will be handled in parts (2) and (3).

For part (2), suppose that $\operatorname{rref}(A)$ does not have a pivot in the last column, and every column of $\operatorname{rref}(C)$ has a pivot. By Exercise 1.1 observe that we must have $m \geq n$, and so $\operatorname{rref}(A)$ is of the form

$$\left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & a_1 \\ 0 & 1 & \cdots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & 1 & a_n \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{array}\right),$$

where the bold-face 0's indicate that the remaining $n - m$ rows are rows of zeros. Hence, our system has one solution, given by $(x_1, \ldots, x_n) = (a_1, \ldots, a_n)$. Conversely, if our system of linear equations has exactly one solution $(a_1, \ldots a_n)$ then our system is equivalent to

$$\begin{cases} x_1 = a_1 \\ x_2 = a_2 \\ \quad \vdots \\ x_n = a_n \end{cases}$$

and so

$$A \sim \left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & a_1 \\ 0 & 1 & \cdots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_n \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{array}\right).$$

Furthermore, noting that the elementary row operations preserve columns, we get

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

Since the matrices above are in reduced row echelon form, then $\operatorname{rref}(A)$ and $\operatorname{rref}(C)$ are as needed.

Finally, for part (3), suppose that $\operatorname{rref}(A)$ does not have a pivot in the last column, and $\operatorname{rref}(C)$ has a column without a pivot. Then, either the variable x_n is free, or there must be two consecutive rows in $\operatorname{rref}(A)$ of the form

$$\left(\begin{array}{ccccccccc|c} 0 & \cdots & \boxed{1} & \cdots & \boldsymbol{*} & \cdots & * & \cdots & & b_i \\ 0 & \cdots & 0 & \cdots & \mathbf{0} & \cdots & \boxed{1} & \cdots & & b_{i+1} \end{array}\right),$$

where the bold-face 0 indicates possibly more than one zero entry in this row. Supposing the jth column has no pivot, we see using back substitution that the variable x_j can be taken to be any real number, which gives a system with infinitely many solutions, as needed. Conversely, if our system has infinitely many solutions, then one of our variables must be free, which yields two consecutive rows in $\operatorname{rref}(A)$ as above. $\square$

Since every matrix either does or does not have a pivot in the last column, we obtain the following corollary.

Corollary 1.27. Every system of linear equations either has no solutions, exactly one solution, or infinitely many solutions.

1.7. Intersections of Lines and Planes: The Row Picture

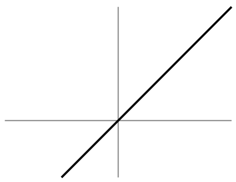
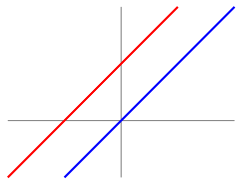
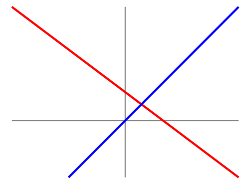
So far, we've been working with systems of linear equations completely algebraically. Let's start looking at some of the geometric implications of what we've learned so far. Note that a linear equation in two variables defines a line in $\mathbb{R}^2$, and a linear equation in three variables

$$ax + by + cz = d$$

defines a plane in $\mathbb{R}^3$. This leads to the following observation.

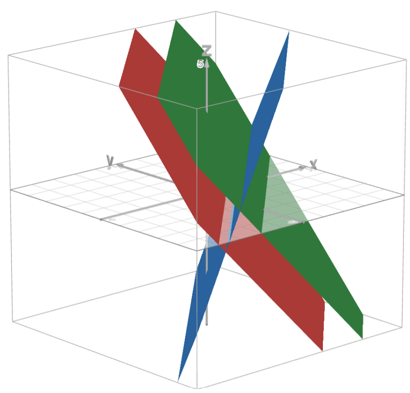
Observation 1.28. Solutions to systems of linear equations in two variables corresponds to intersection points of lines in $\mathbb{R}^2$. Similarly, solutions to systems of linear equations in three variables corresponds to intersection points of planes in $\mathbb{R}^3$.

Let's look first at the two dimensional case. Note that two lines in $\mathbb{R}^2$ can either be the same line, distinct and parallel, or distinct and not parallel

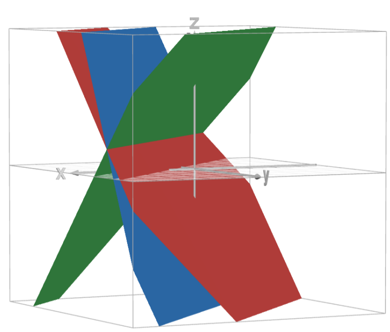




We see geometrically that these cases correspond to systems of linear equations which have infinitely many solutions, no solutions, or exactly one solution.

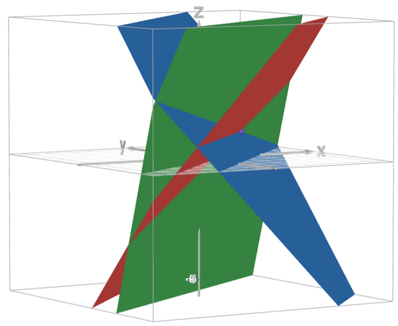
Let's also look at the geometry of systems in three variables. Suppose that we have three linear equations in three variables that correspond to three *distinct* planes in $\mathbb{R}^3$. In this case, we could have two of the planes being distinct and parallel, as in the picture below



or no two of our planes are parallel, in which case our planes can intersect at a line



or at a point



Again we see geometrically that these cases correspond to systems of linear equations which have no solutions, infinitely many solutions, or exactly one solution, which matches what we found in our corollary to Rouché-Capelli.

Exercises

P1.1 Suppose that A is an $m \times n$ matrix. Show that $\mathrm{rref}(A)$ can have *at most* $\max\{m, n\}$ pivots, where $\max\{m, n\}$ is the largest of m and n; that is,

$$\max\{m,n\} = \begin{cases} m & \text{if } m \geq n \\ n & \text{if } m \leq n. \end{cases}$$

P1.2 List all possible reduced row echelon forms of a 3 x 4 matrix with a pivot in exactly two rows. Use '$*$' to denote entries which can be equal to any real number. Make sure to justify how you know you've checked all possible cases.

P1.3 We call a system of linear equations HOMOGENEOUS if the constant term in each equation is zero. For example, the system

$$2x + 3y - z = 0$$
$$-x + 5y + 2z = 0$$

is homogeneous.

(a) Is it possible for a homogeneous system of linear equations to have no solutions? If so, provide an example. If not, provide reasoning.

(b) Is it possible for a homogeneous system of linear equations to have exactly one solution? If so, provide an example. If not, provide reasoning.

(c) Is it possible for a homogeneous system of linear equations to have infinitely many solutions? If so, provide an example. If not, provide reasoning.

P1.4 Give an example of systems of linear equations satisfying the properties below (you should give a different system for each part). If such a system does not exist, explain why not.

(a) A system of linear equations with a unique solution for which the reduced row echelon form of the augmented matrix of the system has a row of 0's.

(b) A system of linear equations with a unique solution that has fewer equations than variables.
(c) A system of linear equations with a unique solution that has more equations than variables.
(d) A consistent system in two variables whose augmented matrix has exactly three pivot columns.
(e) An inconsistent system whose coefficient matrix has a pivot in every row.

P1.5 Show that if $ad - bc \neq 0$ then the system of linear equations

$$\begin{aligned} ax + by &= r \\ cx + dy &= s \end{aligned}$$

has a unique solution.

Chapter 2

Vectors in Euclidean Space

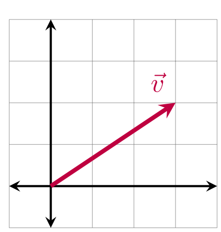
In Section 1.7, we saw that the set of solutions to a system of linear equations in two and three variables corresponds to intersection points of lines and planes in $\mathbb{R}^2$ and $\mathbb{R}^3$. In this chapter, we'll explore an alternate geometric representation of the solution sets by corresponding systems of equations to certain *vector* equations. Let's first develop a notion of vectors in Euclidean space.

2.1. Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$

If someone asks you the question "Where are you?" typically you'd give a response like "I'm near the library, two blocks south of the main entrance". That is, a common way to describe location is to give the displacement (e.g. "two blocks south") from some distinguished point (e.g. "the main entrance of the library"). Furthermore, displacement is described by distance (e.g. "2 blocks") and direction (e.g. "south"). Vectors, our first topic of study in this course, are *the mathematical object we use to represent displacement*. Let's first develop an understanding of vectors in the familiar 2-dimensional Euclidean space, which we denote as $\mathbb{R}^2$.

Definition 2.1. A VECTOR is a mathematical object, typically drawn as an arrow, with a magnitude (i.e. length or distance) and direction.

There are a few ways that we can indicate the magnitude and direction of a vector. One common way to do this is to keep track of the displacement along the axes of a standard coordinate grid. For example, consider the vector in $\mathbb{R}^2$ drawn below



Observe that the displacement this vector represents can be understood by displacement in the x-direction by 3 units and displacement in the y-direction by 2 units. We typically denote this by listing the displacement in each direction, and then wrapping that list up in brackets, as seen below

$$\vec{v} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

This is called the STANDARD COORDINATE REPRESENTATION of $\vec{v}$. Analogously, we could represent displacement in $\mathbb{R}^3$ by keeping track of the displacement in the x, y, and z direction. In general, we have the following.

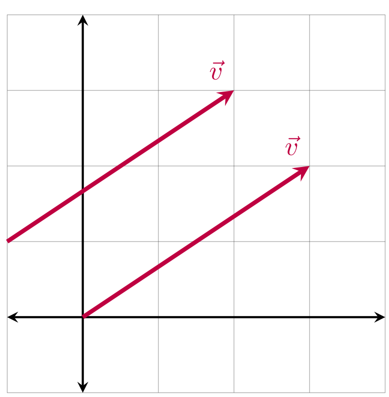
Definition 2.2. Let $\vec{v}$ be the vector in $\mathbb{R}^2$ which represents displacement in the x direction by v_1 units and displacement in the y direction by v_2 units. Then, the STANDARD COORDINATE REPRESENTATION of $\vec{v}$ is given by

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

Similarly, if $\vec{w}$ is a vector in $\mathbb{R}^3$ which represents displacement in the x-direction by w_1, the y-direction by w_2, and the z-direction by w_3, the *standard coordinate representation* of $\vec{w}$ is given by

$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}.$$

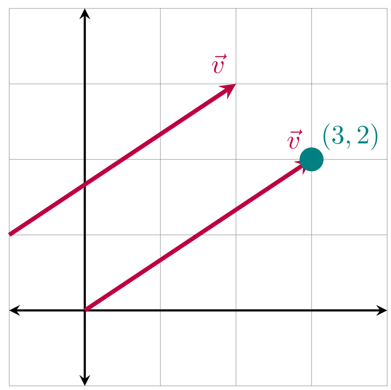
Remark 2.3 (Unrooted Vectors)**.** Note that in the example above, we drew our vector starting from the origin $(0, 0)$. However, it's possible to shift a vector around the plane without changing its direction and magnitude. That is, our vectors are "unrooted", meaning they can "start anywhere." Recall that our vector $\vec{v}$ is the vector which represents displacement in the x direction by 3 and displacement in the y direction by 2, which can be represented as an arrow starting at $(0, 0)$, or at $(-1, 1)$, or anywhere else we'd like



Remark 2.4 (Correspondence Between Vectors and Points)**.** In this course, we'll be studying how certain sets of vectors behave together. One convenient way to do this is to correspond each vector to a *point* in Euclidean space, so that we can think of sets of vectors as familiar geometric objects. Our correspondence will associate

each vector $\begin{pmatrix} x \\ y \end{pmatrix}$ in $\mathbb{R}^2$ to the point (x, y) in $\mathbb{R}^2$.

Graphically, we can visualize this as our vectors corresponding to the point the tip sits at when rooted at the origin. In the example below, we see that the vector $\vec{v}$ corresponds to the point $(3, 2)$, which is simpler to see visually when we root $\vec{v}$ at the origin $(0, 0)$.



> **Activity 2.1.** Suppose that someone gave you the following directions: from your starting point, first walk two blocks west and one block north. After that, walk one block east and then three blocks north. Find the standard coordinate representation of your total displacement.

As we saw in the previous activity, following a sequence of displacement instructions can be done "component-wise". Let's capture this idea algebraically.

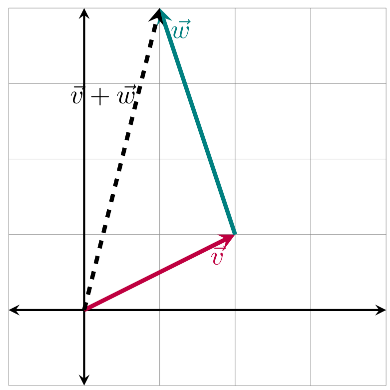
Definition 2.5. Let $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ and $\vec{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ be vectors in $\mathbb{R}^2$ and let c be a scalar (that is, c is a real number). We define the following

$$\vec{v} + \vec{w} := \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \end{pmatrix}, \text{ and } c\vec{v} = \begin{pmatrix} cv_1 \\ cv_2 \end{pmatrix}.$$

Similarly, if $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ and $\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$ are vectors in $\mathbb{R}^3$ we define

$$\vec{v} + \vec{w} := \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ v_3 + w_3 \end{pmatrix}, \text{ and } c\vec{v} = \begin{pmatrix} cv_1 \\ cv_2 \\ cv_3 \end{pmatrix}.$$

Geometrically, we can visualize vector addition by "stacking" vectors head to tail, as in the example below



2.2. Higher Dimensional Vectors

We can generalize our definitions from the previous section to "n-dimensional Euclidean space". Note that this generalization doesn't have any tangible geometric meaning when $n \geq 4$ (at least for those of us that exist in 3-dimensional space). However, many applications of linear algebra use this generalization to imagine a geometry in higher dimensions. So, if you can translate your object of study (which may have no geometric meaning at all) to one in linear algebra, you can often use geometric tools to solve it. This is a powerful and really cool technique, don't let that be lost on you.

Definition 2.6. An n-DIMENSIONAL VECTOR is a list of n real numbers in a specified order, which we'll write in the form

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

We call the real numbers v_i the STANDARD COORDINATES of the vector $\vec{v}$. We denote the set of all n-dimensional vectors by $\mathbb{R}^n$.

We can define our operations on n-dimensional vectors just as we did in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Definition 2.7. Let $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$ and $\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$ be vectors in $\mathbb{R}^n$ and let c be a scalar. We define the following

$$\vec{v} + \vec{w} := \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix}, \text{ and } c\vec{v} = \begin{pmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{pmatrix}.$$

2.3. Vector Equations

The following activity will help us develop a new and important geometric understanding for systems of linear equations.

Activity 2.2 (The Magic Carpet Ride[a]). You are the recipient of a new scholarship at the University of Toronto. As part of your scholarship, you have been gifted two forms of transportation to help you navigate the city: a hover board and a magic carpet. Upon reading the instructions, you find that your two modes of transportation have restrictions on how they operate:

- The hoverboard can only move along the vector $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$

  By this we mean that if the hover board traveled "forward" for one hour, it would move along a "diagonal" path that would result in a displacement of 3 miles East and 1 mile North of its starting location.

- The magic carpet can only move along the vector $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

  By this we mean that if the magic carpet traveled "backward" for one hour, it would move along a "diagonal" path that would result in a displacement of 1 mile West and 2 miles South of its starting location.

Scenario One: Your friend Cramer suggests that you should go visit his friend Gauss outside of Toronto. Cramer tells you that Gauss lives in a cabin that's 107 miles East and 64 miles North of your home here in Toronto.

(1) Is it possible to reach Gauss's cabin using only the hoverboard? If so, how? If not, why not?
(2) Is it possible to reach Gauss's cabin using only the magic carpet? If so how? If not, why not?
(3) Is it possible to reach Gauss's cabin using both the hoverboard and magic carpet? If so how? If not, why not?

[a] Adapted from IOLA

In Activity 2.2, we were able to solve the given vector equations by corresponding them to a system of linear equations. Note that for each system of linear equations, we also have a corresponding vector equation.

Example 2.8. The system of linear equations

$$\begin{cases} x - y - z = 1 \\ 2x - 3y - z = 3 \\ -x + y - z = -3 \end{cases}$$

has the same solution set as the vector equation

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

We summarize this observation below.

Proposition 2.9. The vector equation

$$x_1\vec{v}_1+x_2\vec{v}_2+\cdots+x_n\vec{v}_n=\vec{w}$$

has the same solution set as the linear system represented by the augmented matrix

$$\left(\begin{array}{cccc|c}\vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m & \vec{w}\end{array}\right).$$

We omit the proof of this Proposition here, since the proof just involves keeping track of messy notation. Instead, we encourage the reader to convince themselves of this result by checking several examples.

2.4. Linear Combinations and Spans

Let's continue exploring our magic carpet ride problem.

Activity 2.3 (The Magic Carpet Ride, continued[a]). Recall that you have two modes of transportation: a hoverboard and a magic carpet.

- The hoverboard can only move along the vector $\begin{pmatrix}3\\1\end{pmatrix}$
- The magic carpet can only move along the vector $\begin{pmatrix}1\\2\end{pmatrix}$.

Scenario Two: It turns out that Gauss actually does not like visitors at his cabin. After Gauss learns that Cramer shared the location of his cabin, he starts looking for a new place to live so that you don't bother him again.

(1) Is there anywhere Gauss can move so that you cannot reach him using your hoverboard and magic carpet? Describe the places that you can reach using a combination of the hover board and the magic carpet and those you cannot.

(2) Gauss also starts looking into purchasing a hover cabin, so that you can not reach him by ground transport. Unfortunately for him, your modes of transportation can also fly! Reading the instructions more closely, we find
   - In flying mode, your hoverboard can move forward and backward along the vector $\begin{pmatrix}3\\1\\1\end{pmatrix}$
   - In flying mode, your magic carpet can move forward and backward along the vector $\begin{pmatrix}1\\2\\3\end{pmatrix}$.

> Is there a location for Gauss to hide in his hover cabin so that you can't reach him using your hoverboard and magic carpet in flying mode?
>
> [a]Adapted from IOLA

Let's add some terminology to the vector equations we encountered in Activities 2.2 and 2.3.

Definition 2.10. A LINEAR COMBINATION of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ is a vector of the form $\vec{w} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n$ where the $c_1, c_2, \ldots, c_n$ are scalars called the COEFFICIENTS of the linear combination.

In Activity 2.3, we were asking whether Gauss's location vector could be represented as a linear combination of the hoverboard and magic carpet ride vectors. In Activity 2.3, we instead wanted to understand the *set of all* linear combinations of the hoverboard and magic carpet ride vectors. We define the following

Definition 2.11. The SPAN of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ in $\mathbb{R}^m$ is the set

$$\operatorname{Span}(\vec{v}_1, \ldots, \vec{v}_n) = \{c_1\vec{v}_1 + \cdots + c_n\vec{v}_n \mid c_1, c_2, \ldots, c_n \in \mathbb{R}\}.$$

That is $\operatorname{Span}(\vec{v}_1, \ldots, \vec{v}_n)$ is the set of all linear combinations of vectors $\vec{v}_1, \ldots, \vec{v}_n$.

2.5. Linear Dependence and Independence

Note that if $\vec{d}$ is a vector in $\mathbb{R}^2$ or $\mathbb{R}^3$ we have

$$\operatorname{Span}(\vec{d}) = \{t\vec{d} : t \in \mathbb{R}\}.$$

Observe that this set describes a **line** through the origin in the direction of $\vec{d}$. Now, if $\vec{d}_1, \vec{d}_2$ are two *distinct and nonparallel* vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ we have

$$\operatorname{Span}(\vec{d}_1, \vec{d}_2) = \{t\vec{d}_1 + s\vec{d}_2 \mid t, s \in \mathbb{R}\}.$$

Observe that this describes a **plane through the origin** with direction vectors $\vec{d}_1, \vec{d}_2$. Let's investigate how to determine what type of geometric object a Span of more than two vectors can yield.

> **Activity 2.4.** Consider the vectors
>
> $$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 6 \\ 3 \\ 8 \end{pmatrix}, \vec{v}_3 = \begin{pmatrix} 4 \\ 1 \\ 6 \end{pmatrix}$$
>
> in $\mathbb{R}^3$. Is $\operatorname{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ equal to a line, a plane, or all of $\mathbb{R}^3$?

In Activity 2.4, we were able to argue that $\operatorname{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ was not equal to $\mathbb{R}^3$ by noticing that $\vec{v}_3$ is a linear combination of $\vec{v}_1$ and $\vec{v}_2$. Using this, we were able to see that $\operatorname{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3) = \operatorname{Span}(\vec{v}_1, \vec{v}_2)$ and so the vector $\vec{v}_3$ was "redundant" information. That is, it was not necessary to include the vector $\vec{v}_3$ in our definition of the span, since it already was an element of that set. Finding a "minimal" subset of vectors that generates a span will be an important problem for us in this course.

In this section, we develop some machinery to detect when we have redundant information in our definition of a span.

Definition 2.12 (Geometric definition of linear dependence)**.** We say that vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ are LINEARLY DEPENDENT if for at least one i we have

$$\vec{v}_i \in \operatorname{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_n).$$

Otherwise, the vectors are called LINEARLY INDEPENDENT.

In Activity 2.4, we were able to detect whether small sets of vectors were linearly dependent or independent. When we have three or more independent vectors in our set, checking this by hand can become computationally challenging. Let's develop a more convenient method to detect this.

Definition 2.13 (Algebraic definition of linear dependence)**.** Vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ are LINEARLY DEPENDENT if there exists a "nontrivial" solution (that is, a solution other than $c_1 = c_2 = \cdots = 0$) to the vector equation

$$c_1\vec{v}_1 + \cdots + c_n\vec{v}_n = \vec{0}.$$

Otherwise, the vectors are LINEARLY INDEPENDENT.

Activity 2.5. Show that our two definitions of linear dependence agree. That is,

(1) Explain how the geometric definition of linear dependence implies the algebraic one.
(2) Explain how the algebraic definition of linear dependence implies the geometric one.

Example 2.14. Let's use our new definition to show the vectors

$$\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix}$$

are linearly dependent. To see this, we need to consider solutions to the vector equation

$$x\begin{pmatrix}1\\1\end{pmatrix} + y\begin{pmatrix}1\\0\end{pmatrix} + z\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}. \tag{2.1}$$

Observe that the solutions to this equation are identical to the solutions to the system of linear equations with augmented matrix

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

which has reduced row echelon form

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

Note that this matrix has a column without a pivot, and so by Rouché-Capelli (Theorem 1.26) we know that there are infinitely many solutions to Equation 2.1.

Namely, there has to be a solution other than $(0, 0, 0)$, and so these vectors are linearly dependent.

Example 2.15. Let's use our new definition to show that the vectors

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

are linearly independent. We can set this up similar to the previous problem. Since the matrix

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

has reduced row echelon form

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

which has a pivot in every column, except for the last one, then by Rouché-Capelli (Theorem 1.26) the vector equation

$$x\begin{pmatrix}1\\1\\0\end{pmatrix} + y\begin{pmatrix}3\\1\\2\end{pmatrix} + z\begin{pmatrix}-1\\2\\0\end{pmatrix} = \vec{0}$$

only has **one solution**, namely the trivial solution $x = y = z = 0$. So our vectors are linearly independent, by our algebraic definition of linear independence.

Remark 2.16. Observe that, since the right-hand side of the vector equation

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{0} \tag{2.2}$$

is equal to zero, the last column in augmented matrix corresponding to this equation is going to always remain a column of zeros under any of our elementary row operations. So, it's enough to only consider the pivot columns of the matrix

$$\begin{pmatrix}\vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n\end{pmatrix}.$$

The following observations give us a quick method to detect linear dependence and independence.

(1) If the reduced row echelon form of the matrix $\begin{pmatrix}\vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n\end{pmatrix}$ has a column without a pivot, then by Rouché-Capelli (Theorem 1.26) we know that the system represented by this matrix has infinitely many solutions. Hence, equation (2.2) has a nontrivial solution, and so our vectors are linearly dependent.

(2) If every column of the matrix $\begin{pmatrix}\vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n\end{pmatrix}$ is a pivot column, then the vector equation (2.2) only has one solution, given by the trivial solution $c_1 = c_2 = \cdots = c_n = 0$. Hence, our vectors are linearly independent.

Example 2.17. Determine which of the following sets of vectors are linearly dependent and which are linearly independent. For those that are linearly dependent, find a nontrivial linear combination of the vectors that is equal to $\vec{0}$.

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

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

(3) $\begin{pmatrix}0\\1\\2\\1\end{pmatrix}, \begin{pmatrix}-1\\1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\\1\\0\end{pmatrix}$

(4) Any set of four vectors in $\mathbb{R}^3$ (give a geometric argument and an algebraic argument for your answer)

Solution for (1): We have

$$\begin{pmatrix}1 & -1 & 1\\1 & 1 & 1\\2 & 1 & 1\\1 & 0 & 1\end{pmatrix} \sim \begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\\0 & 0 & 0\end{pmatrix},$$

and so the vector equation

$$x_1\begin{pmatrix}0\\1\\2\\1\end{pmatrix} + x_2\begin{pmatrix}-1\\1\\1\\0\end{pmatrix} + x_3\begin{pmatrix}1\\1\\1\\1\end{pmatrix} = \vec{0}$$

only has one solution $(0, 0, 0)$. Hence, using our algebraic definition, these vectors are $\boxed{\text{linearly independent}}$.

Solution for (2): We have

$$\begin{pmatrix}0 & -1 & 1 & -2\\1 & 1 & 1 & 1\\2 & 1 & 1 & 2\\1 & 0 & 1 & 0\end{pmatrix} \sim \begin{pmatrix}1 & 0 & 0 & 1\\0 & 1 & 0 & 1\\0 & 0 & 1 & -1\\0 & 0 & 0 & 0\end{pmatrix},$$

which has a column without a pivot, and so by Rouché-Capelli (Theorem 1.26) the vector equation

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

has infinitely many solutions. Namely, there must exist a solution other than the trivial one $(0, 0, 0, 0)$. Hence, using our algebraic definition, these vectors are $\boxed{\text{linearly dependent}}$.

Solution for (3): We have

$$\begin{pmatrix} 0 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 2 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and so the vector equation

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

only has one solution $(0, 0, 0, 0)$. Hence, using our algebraic definition, these vectors are $\boxed{\text{linearly independent}}$

Solution for (4): Suppose that we have four vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ in $\mathbb{R}^3$. Then, the vector equation

$$x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 + x_4\vec{v}_4 = \vec{0}$$

will have the same solution set as the augmented matrix

$$\left(A \;\middle|\; \vec{0} \right)$$

where A is a 3×4 matrix. Since each column of A can have *at most* one pivot, then the matrix A has no more than three pivots. Since A has four columns, it must have a column without a pivot. Since the system $\left(A \;\middle|\; \vec{0} \right)$ cannot have a pivot in he last column (since it's a column of zeros), and there's another column without a pivot, then by Rouché-Capelli (Theorem 1.26) this system has infinitely many solutions. Hence, as we argued in part (2) above, we know that these vectors are $\boxed{\text{linearly dependent}}$.

2.6. The Matrix-Vector Form of a Linear System

Recall in Proposition 2.9 we saw that every vector equation can be represented by a system of linear equations and vice versa. We introduce one more way to represent systems of linear equations, which will be useful as we expand our perspectives throughout the course.

Definition 2.18. Let A be an $n \times m$ matrix and $\vec{x}$ a vector in $\mathbb{R}^m$. Write

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \text{ and } \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

The matrix-vector product $A\vec{x}$ is the vector in $\mathbb{R}^m$ defined by

$$A\vec{x} := x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} + \cdots + x_m \begin{pmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{pmatrix}.$$

Example 2.19. Let

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}, \vec{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \text{ and } \vec{w} = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}.$$

Then,

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

Note that $A\vec{w}$ is undefined, since $\vec{w} \in \mathbb{R}^3$ but A only has two columns. In fact, the vector product $A\vec{u}$ is **only** defined for vectors $\vec{u}$ in $\mathbb{R}^2$.

Example 2.20. The vector equation

$$a \begin{pmatrix} 1 \\ 2 \end{pmatrix} + b \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

can be rewritten in matrix-vector form as

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

Observe that we now have multiple ways to represent and study our original problem of solving systems of linear equations. We should aim to become comfortable with all of the equivalent representations below, as we can learn something using each of these perspectives.

The following four representations each have identical sets of solutions.

Representation 1: Vector equation

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} + \cdots + x_m \begin{pmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

Why? Vector equations arose as our motivating problem for the course, the magic carpet ride problem. Vector equations are often helpful when representing a problem **geometrically**.

Representation 2: Systems of Linear Equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= b_n. \end{aligned}$$

Why? This representation is familiar, and can help us understand our problem **algebraically**.

Representation 2: Augmented matrix

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1m} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2m} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} & b_m \end{array}\right)$$

Why? This representation is convenient for computation. We have a concrete algorithm (Gauss-Jordan elimination) to reduce any augmented matrix to a convenient equivalent form.

Representation 4: Matrix-vector equation

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

Why? In a future chapter, we'll see how matrices can be interpreted as "linear transformations". The representation above shows us that this matrix-vector operation sends a vector in $\mathbb{R}^m$ to another vector in $\mathbb{R}^m$. Keep this idea bookmarked for now.

In the following examples, we demonstrate how to move between these different perspectives.

Example 2.21. Consider the matrix-vector equation

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

Observe that finding a vector solution to this matrix equation is equivalent to finding a solution x, y, z to the system of linear equations

$$x + 2y - z = 2$$
$$2x + y + z = 6.$$

This linear system has augmented matrix

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

which is row equivalent to the matrix in reduced row echelon form

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 10/3 \\ 0 & 1 & -1 & -2/3 \end{array}\right).$$

So all solutions to our system of linear equations above can be written as

$$(x, y, z) = (10/3 - t, -2/3 + t, t).$$

Finding solutions to this system of linear equations is also equivalent to solving the vector equation

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

So, any solution (x, y, z) to our linear system should also satify the vector equation above. For example, if we choose the solution $(x, y, z) = (10/3, -2/3, 0)$ this shows that

$$\begin{pmatrix} 2 \\ 6 \end{pmatrix} = \frac{10}{3} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 2 \\ 1 \end{pmatrix} + 0 \cdot \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

That is, the vector $\begin{pmatrix} 2 \\ 6 \end{pmatrix}$ is in the span of the vectors

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \text{ and } \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

Example 2.22. Consider the matrix-vector equation

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

Since the matrix

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

is row equivalent to

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

and this matrix has a pivot in every row, we know that the matrix-vector equation

$$A\vec{x} = \vec{b}$$

has a solution for any vector $\vec{b}$ in $\mathbb{R}^n$. We also know that

$$\text{Span}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \end{pmatrix} \right) = \mathbb{R}^2$$

and that the system of linear equations

$$x + 2y - z = b_1$$

$$2x + y + 3z = b_2$$

has a solution for any $b_1, b_2 \in \mathbb{R}$.

Exercises

P2.1 True or False: For vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^2$, if $\text{Span}(\vec{u}, \vec{v}) = \mathbb{R}^2$ then

$$\text{Span}(\vec{u}, \vec{u} + \vec{v}) = \mathbb{R}^2.$$

If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P2.2 True or False: For vectors $\vec{u}, \vec{v}, \vec{w}$ in $\mathbb{R}^3$, if $\text{Span}(\vec{u}, \vec{v}, \vec{w}) = \mathbb{R}^3$ then

$$\text{Span}(\vec{u} + \vec{v}, \vec{v} + \vec{w}) = \mathbb{R}^3.$$

If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P2.3 Give an example for each of the following, if possible. If it is not possible to construct an example, explain why not.

(a) Three linearly independent vectors in $\mathbb{R}^4$.
(b) Three linearly dependent vectors in $\mathbb{R}^4$.
(c) Three linearly independent vectors in $\mathbb{R}^2$.
(d) Four linearly independent vectors in $\mathbb{R}^3$.
(e) A 4×3 matrix with linearly independent columns.
(f) A 3×4 matrix with linearly independent columns.

P2.4 True or False: if $m > n$, then any set of vectors $\vec{v}_1, \ldots, \vec{v}_m$ in $\mathbb{R}^n$ are linearly dependent. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P2.5 True or False: If $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ are linearly dependent vectors in $\mathbb{R}^n$, then $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are also linearly dependent. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P2.6 Let $\{\vec{u}, \vec{v}, \vec{w}\}$ be a set of linearly independent vectors in $\mathbb{R}^n$. Determine whether the following sets of vectors are linearly dependent or independent, and justify your answer.

(a) $\{\vec{u} - \vec{v}, \vec{v} - \vec{w}, \vec{w} - \vec{u}\}$
(b) $\{\vec{u} + \vec{v}, \vec{v} + \vec{w}, \vec{w} + \vec{u}\}$

Chapter 3

Vector Subspaces of $\mathbb{R}^n$

3.1. Vector Spaces

Activity 3.1. Consider the following subsets of $\mathbb{R}^3$, which are graphed in desmos. Use geometric reasoning to determine whether the sets could be equal to a span.

(1) link to desmos graph of $\mathcal{S}$
(2) link to desmos graph of $\mathcal{T}$
(3) link to desmos graph of $\mathcal{R}$

In Activity 3.1, we used geometric reasoning to argue that spans in $\mathbb{R}^3$ need to satisfy two key geometric properties. We define the following.

Definition 3.1. A SUBSPACE V of $\mathbb{R}^n$ is any nonempty subset of $\mathbb{R}^n$ that satisfies both of the following properties

(1) V is closed under vector addition; that is for all $\vec{u}, \vec{v} \in V$ we have $\vec{u} + \vec{v} \in V$
(2) V is closed under scalar multiplication; that is, for all $\vec{u} \in V$ and $k \in \mathbb{R}$ we have $k\vec{u} \in V$

Note that a VECTOR SPACE can be defined more generally to be any set V satisfying the properties above, where we allow our scalars k to come from any "field" (which are algebraic objects analogous to $\mathbb{R}$). In this course, we'll stick with vector subspaces of $\mathbb{R}^n$. For convenience, we will often refer to our subspaces as vector spaces, and remember that in this course they will always be subspaces of $\mathbb{R}^n$.

Example 3.2. Observe that the set

$$\mathcal{S} = \left\{ \begin{pmatrix} x \\ 1 \end{pmatrix} : x \in \mathbb{R} \right\}$$

is not a vector space, since for example $\vec{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \in \mathcal{S}$ but $2\vec{v} \notin \mathcal{S}$.

Theorem 3.4 will generalize our geometric observations from Activity 3.1 to n-dimensional space. We first need a lemma.

Lemma 3.3. *If $\mathcal{S}$ is a set of m linearly independent vectors in $\mathbb{R}^n$, then $m \leq n$.*

Proof. We leave this proof as a Chapter Exercise (see Exercise P3.1) $\square$

Theorem 3.4. *A subset V is a vector subspace of $\mathbb{R}^n$ if and only if exists vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ so that $V = \mathit{Span}\left(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\right)$.*

Proof. Suppose that $V = \operatorname{Span}\left(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\right)$ for vectors $\vec{v}_1, \ldots, \vec{v}_m$ in $\mathbb{R}^n$. Observe that $\vec{0} \in V$ since we can write

$$\vec{0} = 0\vec{v}_1 + \cdots + 0\vec{v}_m,$$

and so V is not empty. Furthermore, for $\vec{u}, \vec{v} \in V$ we can write

$$\vec{u} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m$$

$$\vec{v} = d_1\vec{v}_1 + \cdots + d_m\vec{v}_m$$

for scalars $c_i, d_i \in \mathbb{R}$ and so

$$\vec{u} + \vec{v} = (c_1 + d_1)\vec{v}_1 + \cdots + (c_m + d_m)\vec{v}_m \in \operatorname{Span}(\vec{v}_1, \ldots, \vec{v}_m).$$

Finally, for any $k \in \mathbb{R}$ we have

$$a\vec{u} = (ac_1)\vec{v}_1 + \cdots + (ac_m)\vec{v}_m \in \operatorname{Span}(\vec{v}_1, \ldots, \vec{v}_m).$$

So, V is a subspace of $\mathbb{R}^n$.

Conversely, suppose that V is a vector subspace of $\mathbb{R}^n$. Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ be a linearly independent subset of V. Since V is a vector subspace, we have

$$\operatorname{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m) \subseteq V.$$

Now, by Lemma 3.3 we know that $m \leq n$, and so we may choose m to be maximal. Take any $\vec{v} \in V$. If $\vec{v} = \vec{v}_i \in B$ then certainly $\vec{v} \in \operatorname{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m)$ by writing

$$\vec{v} = 0\vec{v}_1 + \cdots + 1 \cdot \vec{v}_i + \cdots + 0\vec{v}_m.$$

So, suppose that $\vec{v} \notin \operatorname{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m)$. Since m is maximal, the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m, \vec{v}\}$ must be linearly **dependent**. So, there exist scalars $c_1, \ldots, c_{m+1} \in \mathbb{R}$ not all equal to zero so that

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m + c_{m+1}\vec{v} = \vec{0}.$$

Furthermore, $c_{m+1} \neq 0$ since the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ is linearly independent. So we have

$$\vec{v} = \frac{1}{c_{m+1}}\left(c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m\right) \in \operatorname{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m).$$

Hence, $V = \operatorname{Span}(\vec{v}_1, \ldots, \vec{v}_m)$, as needed. $\square$

Definition 3.5. If V is a vector space with $V = \operatorname{Span}(\vec{v}_1, \ldots, \vec{v}_m)$, then we call $\{\vec{v}_1, \ldots, \vec{v}_m\}$ a SPANNING SET (or GENERATING SET) for V.

Activity 3.2. Consider the following subset of $\mathbb{R}^3$

$$V = \left\{ \begin{pmatrix} x-y \\ x+y \\ y \end{pmatrix} : x, y \in \mathbb{R} \right\}.$$

(1) Show that V is a vector space by using Definition 3.1.
(2) Show that V is a vector space by finding a generating set for V.

3.2. Bases

Definition 3.6. Let V be a vector subspace of $\mathbb{R}^n$. A subset $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \dots, \vec{b}_m\}$ of $\mathbb{R}^n$ is called a BASIS if it's a linearly independent generating set. That is, the vectors $\vec{b}_1, \dots, \vec{b}_m$ are linearly independent and we have $V = \operatorname{Span}(\vec{b}_1, \dots, \vec{b}_m)$.

Activity 3.3. Determine which of the following sets are bases for $\mathbb{R}^3$

$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \right\}$$

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

$$\mathcal{B}_3 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

$$\mathcal{B}_4 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Remark 3.7. In the example above, we saw that bases are not unique, however each set which was a basis had the same number of elements. This turns out to be true in general.

Theorem 3.8. *Let V be a vector subspace of $\mathbb{R}^n$. Then, the size of any basis for V is unique.*

Proof. Suppose that V is a vector subspace of $\mathbb{R}^n$ with bases

$$\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}, \text{ and } \mathcal{C} = \{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_m\}.$$

We need to show that $k = m$. To do this, we'll use a *proof by contradiction* (see Appendix A.4 for some background on this proof method).

By way of contradiction, suppose that $k \neq m$. Without loss of generality, suppose that $k < m$. For each $i \in \{1, \dots, m\}$ write

$$\vec{u}_i = a_{i1}\vec{v}_1 + a_{i2}\vec{v}_2 + \cdots + a_{ik}\vec{v}_k. \tag{3.1}$$

Now, consider the vector equation

$$\vec{0} = x_1\vec{u}_1 + \cdots + x_m\vec{u}_m \tag{3.2}$$

with unknowns $x_1, \ldots, x_m$. Replacing each $\vec{u}_i$ with its representation in terms of the basis elements from B as in Equation (3.1) and collecting coefficients, we obtain

$$\begin{aligned}\vec{0} = \quad & (x_1a_{11} + x_2a_{21} + \cdots + x_ma_{m1})\vec{v}_1 \\ & + (x_1a_{12} + x_2a_{22} + \cdots + x_ma_{m2})\vec{v}_2 \\ & \vdots \\ & + (x_1a_{1k} + x_2a_{2k} + \cdots + x_ma_{mk})\vec{v}_k.\end{aligned}$$

Now, since the $\vec{v}_i$ are linearly independent, each of the coefficients above must be equal to zero. This yields the system of linear equations

$$\begin{gathered}x_1a_{11} + x_2a_{21} + \cdots + x_ma_{m1} = 0 \\ x_1a_{12} + x_2a_{22} + \cdots + x_ma_{m2} = 0 \\ \vdots \\ x_1a_{k1} + x_2a_{k2} + \cdots + x_ma_{mk} = 0,\end{gathered}$$

which is equivalent to the matrix-vector equation

$$A\vec{x} = \vec{0}$$

where

$$A = \begin{pmatrix} a_{ji} \end{pmatrix} \text{ and } \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}.$$

Since A is a $k \times m$ matrix and $k < m$, there are more columns than rows, which means we have a column with no pivot. Since our system has at least one solution (namely, the solution $\vec{x} = \vec{0}$), then by Rouché-Capelli, this system has infinitely many solutions. Namely, we have a nontrivial solution to (3.2), which means our set $\mathcal{C} = \{\vec{u}_1, \ldots, \vec{u}_m\}$ is linearly dependent, a contradiction. Hence, it must have been that $k = m$. □

This gives rise to the following definition.

Definition 3.9. Let V be a vector subspace of $\mathbb{R}^n$. Then, the DIMENSION of V, denoted $\dim V$, is equal to the size of any basis for V. We define the dimension of the trivial subspace $\{\vec{0}\}$ to be 0.

3.3. Finding Bases

Activity 3.4. Show that $\dim \mathbb{R}^n = n$.

In Activity 3.4 we were able to calculate the dimension of our space by finding a basis for that space (which we called the *standard basis* for $\mathbb{R}^n$). In this section, we derive a method to find a basis for any vector space when we have a generating set.

Activity 3.5. Let $V = \operatorname{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4)$ be a vector subspace of $\mathbb{R}^n$ and

$$A = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 & \vec{v}_4 \end{pmatrix}.$$

Suppose that $\operatorname{rref}(A)$ has a pivot in columns $1, 3$ and 4 but does not have a pivot in column 2.

(1) Show that $\vec{v}_2 \in \operatorname{Span}(\vec{v}_1, \vec{v}_3, \vec{v}_4)$
(2) Show that $\{\vec{v}_1, \vec{v}_3, \vec{v}_4\}$ is a linearly independent set
(3) Conclude that $\dim V = 3$.

The following result generalizes our observations from Activity 3.5.

Lemma 3.10. *Let A be an $n \times m$ matrix of the form*

$$A = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m \end{pmatrix}$$

where the $\vec{v}_i$ are vectors in $\mathbb{R}^n$, and suppose that

$$\operatorname{rref}(A) = \begin{pmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_m \end{pmatrix}.$$

If the column $\vec{x}_m$ of $\operatorname{rref}(A)$ does not have a pivot, then

$$\operatorname{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m) = \operatorname{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{m-1}).$$

In general, we can remove any column of A that is not a pivot column and not change the span of its column vectors.

Proof. Consider the vector equation

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}.$$

Since there is no pivot in the mth column of X, we know that the variable x_m is free. That is, we can set x_m equal to *any real number* and obtain a solution to the vector equation above. Setting $c_m = -1$ gives

$$\vec{v}_m = c_1\vec{v}_1 + \cdots + c_{m-1}\vec{v}_{m-1} \in \operatorname{Span}(\vec{v}_1, \ldots, \vec{v}_{m-1}),$$

as needed. □

By repeated use of Lemma 3.10 we obtain the following.

Theorem 3.11 (Finding Bases)**.** *Let V be the vector subspace of $\mathbb{R}^n$ given by*

$$V = \operatorname{Span}(\vec{v}_1, \ldots, \vec{v}_m).$$

If A is the matrix with column vectors $\vec{v}_1, \ldots, \vec{v}_m$ then the pivot columns of A will form a basis for V. Furthermore, if the reduced row echelon form of A has k pivots, then $\dim(V) = k$.

Exercises

P3.1 Prove Lemma 3.3.

P3.2 Show that a vector space V is nonempty if and only if $\vec{0} \in V$.

P3.3 Show that any set of n linearly independent vectors in $\mathbb{R}^n$ forms a basis for $\mathbb{R}^n$.

P3.4 Determine which of the following sets W is a subspace of $\mathbb{R}^n$ for the indicated value of n. If W is a subspace, provide a proof. If not, explain why not.

(a) $W = \left\{ \begin{pmatrix} 2x+y \\ x-y \\ x+y \end{pmatrix} : x, y \in \mathbb{R} \right\}$ in $\mathbb{R}^3$

(b) $W = \left\{ \begin{pmatrix} x+1 \\ x-1 \end{pmatrix} : x \in \mathbb{R} \right\}$ in $\mathbb{R}^2$

(c) $W = \left\{ \begin{pmatrix} xy \\ xz \\ yz \end{pmatrix} : x, y, z \in \mathbb{R} \right\}$ in $\mathbb{R}^3$

P3.5 Suppose that V and W are both vector subspaces of $\mathbb{R}^n$. Let $V + W$ be the subset of $\mathbb{R}^n$ defined by

$$V + W = \{\vec{v} + \vec{w} \mid \vec{v} \in V \text{ and } \vec{w} \in W\}.$$

Show that $V + W$ is a vector subspace of $\mathbb{R}^n$

P3.6 True or False: if W and V are subspaces of $\mathbb{R}^n$, then

$$W \cap V := \{\vec{x} \in \mathbb{R}^n : \vec{x} \in W \text{ and } \vec{x} \in V\}$$

is a subspace of $\mathbb{R}^n$. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P3.7 True or False: if W and V are subspaces of $\mathbb{R}^n$, then

$$W \cup V := \{\vec{x} \in \mathbb{R}^n : \vec{x} \in W \text{ or } \vec{x} \in V\}$$

is a subspace of $\mathbb{R}^n$. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P3.8 Let V be a vector subspace of $\mathbb{R}^n$ of dimension $1 \le m < n$ and suppose that B is a basis for V. Show that if $\vec{v}$ in $\mathbb{R}^n$ is not an element of V, then the set $B \cup \{\vec{v}\}$ is linearly independent.

P3.9 Let V be a vector subspace of $\mathbb{R}^n$ of dimension $m \ge 1$. Show that any basis B for V can be extended to a basis for $\mathbb{R}^n$. That is, if $B = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_m\}$ show that there exist a basis C for $\mathbb{R}^n$ with $B \subseteq C$.

P3.10 Let V and W be vector subspaces of $\mathbb{R}^5$ with $V \cap W = \{\vec{0}\}$. Suppose that V has basis $\{\vec{v}_1, \vec{v}_2\}$ and W has basis $\{\vec{w}_1, \vec{w}_2\}$. Find a basis for the vector space $V + W$, and justify how you know this is a basis.

P3.11 True or False: If V and W are vector subspaces of $\mathbb{R}^n$, then $\dim(V + W) = \dim(V) + \dim(W)$. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P3.12 Is it possible to find two vector subspaces V and W of $\mathbb{R}^3$ with $V \cap W = \{\vec{0}\}$ so that $\dim V = \dim W = 2$? If it is possible, give an example and justify that your example satisfies these conditions. If it is not possible, explain why not.

Chapter 4

Fundamental Subspaces and the Geometry of Systems

4.1. Three Fundamental Subspaces

Let A be an $m \times n$ matrix. Recall that the matrix-vector product $A\vec{x}$ is defined for vectors $\vec{x} \in \mathbb{R}^n$. Furthermore, for each vector $\vec{x}$ in $\mathbb{R}^n$, the matrix-vector product $A\vec{x}$ yields a vector $\mathbb{R}^m$. Every matrix has three "fundamental subspaces", which will help us gain a further geometric understanding for the solution set to a systems of linear equations, by corresponding our system to a matrix-vector product. We define two of these spaces below.

Definition 4.1. Let A be an $m \times n$ matrix with

$$A = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}.$$

Then, the COLUMN SPACE of A is the subspace of $\mathbb{R}^m$ given by

$$\mathrm{Col}(A) := \mathrm{Span}(\vec{v}_1, \ldots, \vec{v}_n).$$

The NULL SPACE of A is the subspace of $\mathbb{R}^n$ given by

$$\mathrm{Nul}(A) := \{\vec{x} \in \mathbb{R}^m \mid A\vec{x} = \vec{0}\}.$$

Activity 4.1. Let

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}.$$

(1) Find a vector $\vec{v} \in \mathrm{Col}(A)$.
(2) Find a vector $\vec{w} \in \mathrm{Nul}(A)$.

Remark 4.2. Pay attention to the fact that $\mathrm{Col}(A)$ and $\mathrm{Nul}(A)$ are subsets of *different* ambient spaces. When A is $m \times n$ we have

$$\mathrm{Col}(A) \subseteq \mathbb{R}^{\boxed{m}} \text{ and } \mathrm{Nul}(A) \subseteq \mathbb{R}^{\boxed{n}}.$$

> **Activity 4.2.** Show that $\mathrm{Col}(A)$ and $\mathrm{Nul}(A)$ are in fact vector spaces.

The activity above allows us to define the following.

Definition 4.3. The *rank* of a matrix A is given by

$$\mathrm{rank}(A) := \dim \mathrm{Col}(A)$$

and the nullity of A is given by

$$\mathrm{nullity}(A) := \dim \mathrm{Nul}(A).$$

> **Activity 4.3.** Find the rank and nullity of the following matrices.
>
> (1) $A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
>
> (2) $B = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 3 & 2 & -1 \\ 0 & -1 & -1 & 1 \\ 0 & 1 & 1 & -1 \end{pmatrix}$

To define our third fundamental subspace, it will be convenient to define a matrix operation. We have the following.

Definition 4.4. Let A be the $m \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Then, the transpose of A is the $n \times m$ matrix $A^\top$ given by

$$A^\top = \begin{pmatrix} a_{11} & a_{11} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}.$$

That is, $A^\top$ is the matrix with column vectors equal to the rows of A.

Example 4.5. Let

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}.$$

Then

$$A^\top = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

Definition 4.6. The row space of an $m \times n$ matrix A is the vector subspace $\mathrm{Row}(A)$ of $\mathbb{R}^n$ given by

$$\mathrm{Row}(A) = \mathrm{Col}(A^\top).$$

That is, $\mathrm{Row}(A)$ is the subspace of $\mathbb{R}^n$ spanned by the row vectors of A.

Note that the row space of a matrix doesn't have much geometric meaning for us at the moment, but it will later on in the course. So let's bookmark this story for now and come back to this once we've developed a bit more machinery.

4.2. Rank-Nullity

Activity 4.4. Let A be a 3×3 matrix.
(1) Suppose that A has exactly two pivot columns, which are located in columns 1 and 3. Show that $\operatorname{rank}(A) = 2$ and $\operatorname{nullity}(A) = 1$.
(2) Suppose that A has exactly one pivot column, which is located in column 1. Show that $\operatorname{rank}(A) = 1$ and $\operatorname{nullity}(A) = 2$.

The following result generalizes our observations from above.

Theorem 4.7. *Let A be an $m \times n$ matrix with r pivot columns. Then, $rank(A) = r$ and $\operatorname{nullity}(A) = n - r$. That is, the rank of A is equal to the number of pivot columns of A, and the nullity of A is the number of non-pivot columns of A.*

Proof. Suppose that $\operatorname{rref}(A) = \begin{pmatrix} \vec{v}_1 & \cdots \vec{v}_n \end{pmatrix}$. Without loss of generality, suppose that $\vec{v}_1, \ldots, \vec{v}_r$ contain pivots, and $\vec{v}_{r+1}, \ldots, \vec{v}_n$ do not contain pivots. Then $\operatorname{Col}(A)$ has dimension r by Theorem 3.11, as needed.

Next, observe that

$$\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \in \operatorname{Nul}(A) \quad \text{if and only if} \quad x_1\vec{v}_1 + \cdots + x_n\vec{v}_n = \vec{0}.$$

In the vector equation above, $x_{r+1}, \ldots, x_n$ are free, and so we can write

$$\begin{aligned} x_1 &= a_{1,r+1}x_{r+1} + a_{1,r+2}x_{r+2} + \cdots + a_{1,n}x_n \\ x_2 &= a_{2,r+1}x_{r+1} + a_{2,r+2}x_{r+2} + \cdots + a_{2,n}x_n \\ &\vdots \\ x_r &= a_{r,r+1}x_{r+1} + a_{r,r+2}x_{r+2} + \cdots + a_{r,n}x_n. \end{aligned}$$

So, we have

$$\operatorname{Nul}(A) = \operatorname{Span}\left(\begin{pmatrix} a_{1,r+1} \\ a_{2,r+1} \\ \vdots \\ a_{r,r+1} \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} a_{1,r+2} \\ a_{2,r+2} \\ \vdots \\ a_{r,r+2} \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{r,n} \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right).$$

Finally, observe that the $m - r$ vectors above are linearly independent, since the matrix with the vectors above as its columns has a pivot in every column. $\square$

We have the following consequence to Theorem 4.7, often referred to as the RANK-NULLITY THEOREM. Observe that this gives a connection between the geometry of the column space and the geometry of the null space. This connection will be revisited in the next chapter as well.

Corollary 4.8 (The Rank-Nullity Theorem)**.** Let A be an $m \times n$ matrix. Then,

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n.$$

Proof. Suppose that the matrix A has r pivot columns. By Theorem 3.11 we know that $\operatorname{rank}(A) = r$. So the result follows by Theorem 4.7. $\square$

4.3. Homogeneous Systems and the Geometry of Systems

Now that we've developed some machinery, we can return to our problem of describing the solution set to a system of linear equations. We define the following.

Definition 4.9. A system of linear equations is called HOMOGENEOUS if the constant coefficients are all equal to zero. For example, the system

$$\begin{cases} 2x - y = 0 \\ x + 3y = 0 \end{cases}$$

is homogeneous.

On your Webwork assignments, you saw several examples of solution sets to homogeneous systems that could be described as a span. From what we did in the previous sections, we can now prove that this is always the case.

Theorem 4.10. *The solution set to any homogeneous system of equations is a vector space. Furthermore, if the system has coefficient matrix A, then the solution set is equal to* $\operatorname{Nul}(A)$.

Proof. Observe that a homogeneous system can be represented in matrix-vector form as

$$A\vec{x} = \vec{0}.$$

So, the set of solutions to a homogeneous system is equal to $\operatorname{Nul}(A)$, which we showed was a vector space in Activity 4.2. $\square$

Example 4.11. The system

$$\begin{cases} x + 2y + 4z = 0 \\ \quad x + y - z = 0 \\ \qquad y + 5z = 0 \end{cases}$$

has solution set equal to the null space of

$$\begin{pmatrix} 1 & 2 & 4 \\ 1 & 1 & -1 \\ 0 & 1 & 5 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -6 \\ 0 & 1 & 5 \\ 0 & 0 & 0 \end{pmatrix}.$$

So, by Rank-Nullity (Corollary 4.8) we know that the solution set to our system can be described by a line through the origin. We can find this line by calculating the null space: note that $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathrm{Nul}(A)$ if and only if

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

and so if we choose z to be free, all solutions to this equation can be described parametrically as $(x, y, z) = (-6z, -5z, z)$. Hence

$$\mathrm{Nul}(A) = \left\{ \begin{pmatrix} 6z \\ -5z \\ z \end{pmatrix} : z \in \mathbb{R} \right\} = \mathrm{Span}\left(\begin{pmatrix} 6 \\ -5 \\ 1 \end{pmatrix} \right).$$

So, the solution set to our system is equal to the line passing through the origin in the direction of $\begin{pmatrix} 6 \\ -5 \\ 1 \end{pmatrix}$.

Remark 4.12. Theorem 4.10 tells us that we can quickly detect what the solution set to a *homogeneous* system of linear equations looks like. In the following theorem, we extend this to *any* system of linear equations.

Theorem 4.13. *The solution set to a consistent system of linear equations in n variables with coefficient matrix A is equal to*

$$\vec{p} + \mathrm{Nul}(A) = \{\vec{p} + \vec{v} \mid \vec{v} \in \mathrm{Nul}(A)\}$$

where $\vec{p}$ is any particular vector solution to the system of linear equations.

Remark 4.14. Observe that this generalizes Theorem 4.10, since if our system is homogeneous we can take $\vec{p} = \vec{0}$ as a particular solution, which gives

$$\vec{p} + \mathrm{Nul}(A) = \vec{0} + \mathrm{Nul}(A) = \mathrm{Nul}(A).$$

Proof. Suppose that our system of linear equations has matrix-vector representation $A\vec{x} = \vec{b}$ for a vector $\vec{b} \in \mathbb{R}^n$, and suppose that $\vec{y}$ is any solution. Observe that $\vec{y} - \vec{p}$ is a solution to $A\vec{x} = 0$ and so we can write

$$\vec{y} - \vec{p} = \vec{v}$$

for some $\vec{v} \in V$. This gives $\vec{y} = \vec{v} + \vec{p}$ for $\vec{v} \in V$ as desired. Conversely, if $\vec{v} \in V$, it can be shown that $\vec{p} + \vec{v}$ is a solution to $A\vec{x} = \vec{b}$, as needed. □

Remark 4.15. We call the sets $\vec{p} + V$ TRANSLATED VECTOR SPACES (or TRANSLATED SPANS). Geometrically, these sets look like a vector space translated by some fixed vector $\vec{p}$ as in the examples pictured below

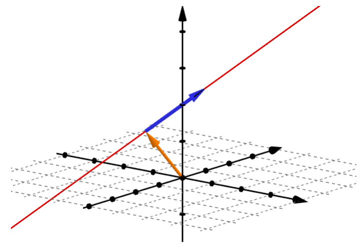


Figure 1. A translated one-dimensional vector space in $\mathbb{R}^3$ is a line

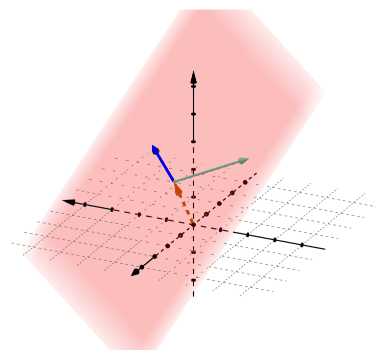


Figure 2. A translated two-dimensional vector space in $\mathbb{R}^3$ is a plane

Activity 4.5. Determine whether the solution set for each of the following systems is empty, a point, a line, or a plane in $\mathbb{R}^3$.

(1) System 1
$$\begin{cases} x+2y+4z=1 \\ x+y-z=2 \\ y+5z=-1 \end{cases}$$

(2) System 2
$$\begin{cases} x+2y+2z=5 \\ x+y+z=0 \\ 3x+3z=1 \end{cases}$$

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

(4) System 2
$$\begin{cases} x+2y+2z=a \\ x+y+z=b \\ 3x+3z=c \end{cases}$$

Note that a, b, and c denote unknown real numbers.

Exercises

P4.1 If the column space of $\begin{pmatrix} 1 & 2 & -1 \\ 1 & 1 & 1 \\ 1 & 2 & c \end{pmatrix}$ has basis $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} \right\}$, what must c be equal to?

P4.2 If the null space of $\begin{pmatrix} 2 & 1 & a \\ 1 & 2 & b \end{pmatrix}$ has basis $\left\{ \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \right\}$, what must a and b be equal to?

P4.3 Find a matrix with at least two rows whose null space has basis $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix} \right\}$.

P4.4 True or False: There exists a 3×3 matrix whose column space and null space have the same dimension. If such a matrix exists, give an example of one and justify why this is an example. If it is not possible to construct this matrix, explain why not.

P4.5 True or False: If A is an $n \times n$ matrix, then $\text{nullity}(A)$ is equal to the number of rows of zeros in $\text{rref}(A)$. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

In the following problems, we define the *sum* of two $m \times n$ matrices A and B to be the matrix formed by taking the sum of the components of A and B. That is, if

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \text{ and } B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}$$

then

$$A + B = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix}.$$

P4.6 True or False: If A and B are $m \times n$ matrices, then

$$\text{rank}(A + B) = \text{rank}(A) + \text{rank}(B).$$

If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P4.7 Suppose that A and B are $n \times n$ matrices with $A\vec{x} \neq B\vec{x}$ for any nonzero $\vec{x}$ in $\mathbb{R}^n$. Find the rank of $A - B$.

Chapter 5

Linear Transformations

5.1. Linearity

Our next main object of study this semester is *linear transformations.* Geometricaly, these are transformations which leave "flat spaces flat".

Definition 5.1. A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is called LINEAR if it satisfies the following two properties:

(1) $F(\vec{x} + \vec{y}) = F(\vec{x}) + F(\vec{y})$, and
(2) $F(c\vec{x}) = cF(\vec{x})$

for any vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

Note that we'll often use the word *transformation* instead of the word *function* in this class, to emphasize that we will be studying how these functions "transform" space.

Activity 5.1. Determine which of the following are linear transformations

(1) F: $\mathbb{R}^2 \to \mathbb{R}^2$ defined by $F(\vec{x}) = 2\vec{x}$

(2) $G : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$G\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x^2 \\ y^2 \end{pmatrix}$$

(3) $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$T\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x+1 \\ y-1 \end{pmatrix}.$$

In the next activity, we'll see that linear transformations "send lines to lines". You'll generalize this activity in P5.2 to show that linear transformations "keep flat spaces flat".

Activity 5.2. Suppose that $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation. Recall that a line in $\mathbb{R}^2$ can be described by the set

$$S = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 : y = mx + b \right\}$$

for some fixed real numbers m and b. Show that the set

$$F(S) := \{F(\vec{v}) : \vec{v} \in S\}$$

is either a line or a point.

5.2. Matrix Transformations

Let A be an $m \times n$ matrix. Recall that the matrix-vector product $A\vec{x}$ is defined for a vector $\vec{x}$ in $\mathbb{R}^n$ and yields a vector $\vec{y} = A\vec{x}$ in $\mathbb{R}^m$. So, this product defines a function, which we'll give a special name and notation to.

Definition 5.2. Let A be an $m \times n$ matrix. Then, the MATRIX TRANSFORMATION associated to A is the function $T_A : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$T_A(\vec{x}) := A\vec{x}.$$

We have the following observation.

Proposition 5.3. Every matrix transformation is a linear transformation.

Proof. Let A be an $m \times n$ matrix, and write $A = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}$. Take any vectors

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ and } \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

in $\mathbb{R}^n$. Then we have

$$\begin{aligned} T_A(\vec{x} + \vec{y}) &= A(\vec{x} + \vec{y}) \\ &= A \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} \\ &= (x_1 + y_1)\vec{v}_1 + (x_2 + y_2)\vec{v}_2 + \cdots + (x_n + y_n)\vec{v}_n \\ &= (x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n) + (y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_n\vec{v}_n) \\ &= A\vec{x} + A\vec{y} \\ &= T_A(\vec{x}) + T_A(\vec{y}). \end{aligned}$$

Now, take any $c \in \mathbb{R}$. Then we have

$$\begin{aligned}
T_A(c\vec{x}) &= A(c\vec{x}) \\
&= A\begin{pmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{pmatrix} \\
&= cx_1\vec{v}_1 + cx_2\vec{v}_2 + \cdots + cx_n\vec{v}_n \\
&= c(x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n) \\
&= c(A\vec{x}) \\
&= cT_A(\vec{x}),
\end{aligned}$$

as needed. □

Activity 5.3. Suppose that $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation given by

$$F(\vec{e}_1) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ and } F(\vec{e}_2) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

(1) Find $F\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$ and $F\left(\begin{pmatrix} 2 \\ 3 \end{pmatrix}\right)$.

(2) Find a formula for $F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right)$.

(3) Find a 2×2 matrix A so that $F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = A\begin{pmatrix} x \\ y \end{pmatrix}$.

The previous activity generalizes.

Theorem 5.4. *Every linear transformation is a matrix transformation. In particular, if $F : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then $F = T_A$ where*

$$A = \begin{pmatrix} F(\vec{e}_1) & F(\vec{e}_2) & \cdots & F(\vec{e}_n) \end{pmatrix}.$$

Note that the vectors $\vec{e}_1, \ldots, \vec{e}_m$ denote the standard basis for $\mathbb{R}^n$, given by

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \vec{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

Proof. Suppose that $F : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and let A be defined as above. Observe that we have

$$F(\vec{e}_i) = A\vec{e}_i \tag{5.1}$$

for every $i \in \{1, \ldots, n\}$. Now, take any $\vec{x} \in \mathbb{R}^m$. Since $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ forms a basis for $\mathbb{R}^m$ we can write $\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n$. So, we have

$$\begin{aligned}
F(\vec{x}) &= F(x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n) \\
&= x_1F(\vec{e}_1) + x_2F(\vec{e}_2) + \cdots + x_nF(\vec{e}_n) && \text{by linearity of } F \\
&= x_1A\vec{e}_1 + x_2A\vec{e}_2 + \cdots + x_nA\vec{e}_n && \text{by Equation (5.1)} \\
&= A(x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n) && \text{by linearity of } T_A \\
&= T_A(\vec{x}).
\end{aligned}$$

So, we have $F(\vec{x}) = T_A(\vec{x})$ for every vector $\vec{x}$ in $\mathbb{R}^n$, which gives $F = T_A$. □

Definition 5.5. Given a matrix A, we call T_A the MATRIX TRANSFORMATION corresponding to the matrix A. Given a linear transformation $F : \mathbb{R}^n \to \mathbb{R}^m$, we call the $m \times n$ matrix A_F constructed in Theorem 5.4 the DEFINING MATRIX of the transformation F.

Activity 5.4. Find the defining matrices for the following linear transformations.

(1) $F : \mathbb{R}^3 \to \mathbb{R}^3$, $\vec{x} \mapsto 2\vec{x}$

(2) $G : \mathbb{R}^2 \to \mathbb{R}^3$ which rotates every vector 90° counterclockwise about the origin.

Activity 5.5. Recall from Activity 5.1 that the function $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$T\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x+1 \\ y-1 \end{pmatrix}$$

is not linear. Determine what's wrong with the following argument.

We have

$$T(\vec{e}_1) = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, T(\vec{e}_2) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \text{and } T(\vec{e}_3) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

So, $T(\vec{x}) = A\vec{x}$ where

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

By Proposition 5.3, every matrix transformation is linear, and so T is a linear transformation.

Activity 5.6. Show that the function $H : \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$H\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+y \\ x \\ 0 \end{pmatrix}$$

is linear by showing that it's equal to a matrix transformation.

5.3. Function Composition and the Matrix Product

Given sets A, B, and C and functions $f : A \to B$ and $g : B \to C$, recall that the composite function $f \circ g : A \to C$ is defined by

$$(g \circ f)(a) = g(f(a)).$$

Let's look at how function composition behaves with linear transformations. Given an $m \times k$ matrix A and an $k \times n$ matrix B, we have the associated linear transformations

$$T_A : \mathbb{R}^k \to \mathbb{R}^m, \vec{x} \mapsto A\vec{x}$$

$$T_B : \mathbb{R}^n \to \mathbb{R}^k, \vec{x} \mapsto B\vec{x}.$$

Then $T_A \circ T_B : \mathbb{R}^n \to \mathbb{R}^m$ is the function defined by

$$(T_A \circ T_B)(\vec{x}) = A(B\vec{x}). \tag{5.2}$$

In Chapter Exercise 5.3, you'll show that the composition of linear functions is linear. So, by Theorem 5.4, there exists an $m \times n$ matrix C so that

$$T_A \circ T_B = T_C.$$

We'll call this matrix the product of A and B.

Definition 5.6. Let A be an $m \times k$ matrix and B a $k \times n$ matrix. Then, the MATRIX PRODUCT of A and B is the $m \times n$ matrix C satisfying

$$T_A \circ T_B = T_C.$$

We write $C = AB$.

Let's develop a method to calculate the matrix product AB. Write

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

and suppose $B = \begin{pmatrix} \vec{b}_1 & \cdots & \vec{b}_n \end{pmatrix}$. Then we have

$$\begin{aligned} A(B\vec{x}) &= A\left(x_1\vec{b}_1 + x_2\vec{b}_2 + \cdots + x_n\vec{b}_n\right) \\ &= x_1A\vec{b}_1 + x_2A\vec{b}_2 + \cdots + x_nA\vec{b}_n \\ &= C\vec{x}, \end{aligned}$$

where $C = \begin{pmatrix} A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_n \end{pmatrix}$. Since the defining matrix of a linear transformation is unique, we must have

$$\boxed{AB = \begin{pmatrix} A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_m \end{pmatrix}}.$$

Let's look at some examples.

Example 5.7. Let

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

Since A is 2×3 and B is 3×2, the matrix product AB is a 2×2 matrix. We have

$$AB = \begin{pmatrix} \vec{a}_1 & \vec{a}_2 \end{pmatrix}$$

where

$$\vec{a}_1 = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \text{ and } \vec{a}_2 = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}.$$

So,

$$AB = \begin{pmatrix} -2 & 5 \\ 2 & 4 \end{pmatrix},$$

With practice, we can perform this computation a bit more quickly. Let's perform the steps above by just keeping track of how we're generating each entry:

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

$$\begin{pmatrix} \color{red}{1} & \color{red}{0} & \color{red}{3} \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & \color{red}{2} \\ 0 & \color{red}{3} \\ -1 & \color{red}{1} \end{pmatrix} = \begin{pmatrix} -2 & \color{red}{5} \\ & \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 3 \\ \color{red}{1} & \color{red}{1} & \color{red}{-1} \end{pmatrix} \begin{pmatrix} \color{red}{1} & 2 \\ \color{red}{0} & 3 \\ \color{red}{-1} & 1 \end{pmatrix} = \begin{pmatrix} -2 & 5 \\ \color{red}{2} & \end{pmatrix}$$

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

Remark 5.8. Note that if A is an $m \times k$ matrix and B is $\ell \times n$ matrix, the matrix product AB is only defined when $k = \ell$. So, in the example above, the matrix BA also happens to be defined and gives a 3×3 matrix.

Let's practice a few more examples.

Example 5.9. Let

$$A = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 2 & 3 & 1 \end{pmatrix}, B = \begin{pmatrix} 0 & -1 \\ 3 & 0 \\ 1 & 0 \\ -1 & 4 \end{pmatrix} \text{ and } C = \begin{pmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & -2 & 1 \end{pmatrix}.$$

Then,

$$AB = \begin{pmatrix} 2 & -5 \\ 8 & 4 \end{pmatrix}, BA = \begin{pmatrix} 0 & -2 & -3 & -1 \\ 3 & 0 & 3 & -3 \\ 1 & 0 & 1 & -1 \\ -1 & 8 & 11 & 5 \end{pmatrix}, \text{ and } CB = \begin{pmatrix} 5 & -1 \\ 6 & 4 \\ -3 & 4 \end{pmatrix}.$$

Note that the matrix products BC, AC, and CA are undefined, because they do not have compatible dimensions.

5.4. Geometric Rank-Nullity

Recall, for a function $F : A \to B$, we call A the DOMAIN and B the CODOMAIN. We define two more sets that will help important for us to understand linear transformations.

Definition 5.10. Given a function $F : \mathbb{R}^n \to \mathbb{R}^m$, the KERNEL of F is the subset of $\mathbb{R}^n$ given by

$$\ker(F) := \{\vec{x} \in \mathbb{R}^n \mid F(\vec{x}) = \vec{0}\}.$$

The IMAGE of F is the subset of $\mathbb{R}^m$ given by

$$\operatorname{im}(F) := \{\vec{y} \in \mathbb{R}^m \mid F(\vec{x}) = \vec{y} \text{ for some } \vec{x} \in \mathbb{R}^n\}.$$

Activity 5.7. Consider the function

$$F : \mathbb{R}^3 \to \mathbb{R}^2, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x - y \\ z \end{pmatrix}.$$

Find a vector $\vec{x}$ in $\ker(F)$. Then, find a vector $\vec{y}$ in $\operatorname{im}(F)$.

Activity 5.8. Show that $\ker(F) = \operatorname{Nul}(A_F)$ and $\operatorname{im}(F) = \operatorname{Col}(A_F)$.

The previous lecture activity, along with our work in Chapter 4, allows us to define the following.

Definition 5.11. The RANK of a linear transformation F is the dimension of $\operatorname{im}(F)$. The NULLITY of a linear transformation F is the dimension of $\ker(F)$.

Activity 5.9. Find the rank and nullity of the linear transformation from Activity 5.7.

Proposition 5.12 (Geometric Rank-Nullity)**.** Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then, $n = \dim(\ker(F)) + \dim(\operatorname{im}(F))$.

Proof. This follows by Rank-Nullity (Corollary 4.8) along with Activity 5.8. □

The following example will illustrate a geometric interpretation of rank-nullity. Note that you will not be tested on anything from the rest of this section, since it's a bit beyond the scope of this course. Also note that I'm not going to be careful with a new definition here, my hope is that this example helps you build some geometric intuition.

Example 5.13. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation with defining matrix

$$F = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Observe that

$$\ker(F) = \operatorname{Span}\left(\begin{pmatrix}1\\1\end{pmatrix}\right).$$

This tells us that under the function F, everything on the line $x = y$ gets "identified" with $\vec{0}$. Let's see what else gets identified under this map. We have

$$\begin{aligned} F(\vec{x}) &= F(\vec{y}) \\ \Leftrightarrow F(\vec{x}) - F(\vec{y}) &= \vec{0} \\ \Leftrightarrow F(\vec{x} - \vec{y}) &= \vec{0} \\ \Leftrightarrow \vec{x} - \vec{y} &\in \ker(F). \end{aligned}$$

That is, the vectors $\vec{x}$ and $\vec{y}$ will be identified under F if and only if

$$\vec{x} = \vec{y} + (\text{something in the kernel}).$$

So, the line passing through $\vec{x}$ that's parallel to $\ker(F)$ all gets collapsed onto one single point. Once you perform all of these identifications, what you end up with is the image. Here's a clip from a 3blue1brown video for what this identification looks like.

In general, if $F : \mathbb{R}^m \to \mathbb{R}^n$ is a linear map, we get

$$\mathbb{R}^m / \ker(F) \cong \operatorname{im}(F)$$

where the thing on the left-hand side is called a "quotient", and we should think of it as the "smooshing down" we saw when we identified everything in the kernel (and then everything else in $\mathbb{R}^2$ that also needed to be identified). The equation above is an example of something called the "first isomorphism theorem" that you would see in an abstract algebra course.

To see what this has to do with the dimension, observe that when we identify vectors in the kernel, we "lose" the dimension of the kernel. That is,

$$\dim(\mathbb{R}^m / \ker(F)) = \dim(\mathbb{R}^m) - \dim(\ker(F)).$$

Since the quotient gives us something that looks like the image, we obtain the equality

$$\dim(\mathbb{R}^m) - \dim(\ker(F)) = \dim(\operatorname{im}(F)).$$

Noting that $\dim(\mathbb{R}^m) = m$ and rearranging the equation above precisely gives the geometric version of the rank-nullity theorem (Proposition 5.12).

Exercises

P5.1 Let F be a linear transformation. Show that $F(\vec{0}) = \vec{0}$.

P5.2 Let V be a vector subspace of $\mathbb{R}^n$ and suppose that $F : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Show that

$$F(V) := \{f(v) \mid v \in V\}$$

is a vector subspace of $\mathbb{R}^m$. (This captures our intuitive definition of linear transformations as functions that send "flat spaces to flat spaces").

P5.3 Show that the composition of linear functions is linear. That is, if

$$F : \mathbb{R}^n \to \mathbb{R}^k \text{ and } G : \mathbb{R}^k \to \mathbb{R}^m$$

are linear functions, show that $G \circ F : \mathbb{R}^n \to \mathbb{R}^m$ is also linear.

P5.4 True or False: If A^2 is equal to the zero matrix, then A is equal to the zero matrix. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P5.5 A matrix is called NILPOTENT if there exists a positive integer m so that A^m is equal to the zero matrix. True or False: if A and B are both nilpotent, then AB is also nilpotent. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P5.6 Let A and B be $n \times n$ matrices. Show that $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$.

Note that the following problems should have been placed in Chapter 6! Apologies for the confusion

---

P5.7 True or False: if a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ is one-to-one, then it is also onto. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P5.8 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. If $n < m$, show that F cannot be surjective.

P5.9 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. If $n > m$, show that F cannot be injective.

P5.10 Let $F : V \to W$ be a linear transformation between vector spaces V and W. Show that F is injective if and only if $\ker(F) = \{\vec{0}\}$.

P5.11 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and $\{\vec{v}_1, \ldots, \vec{v}_n\}$ a basis for $\mathbb{R}^n$. Show that F is injective if and only if the set $\{F(\vec{v}_1), \ldots, F(\vec{v}_n)\}$ is linearly independent.

P5.12 Let $F : \mathbb{R}^7 \to \mathbb{R}^3$ be a surjective linear transformation. Find the dimension of $\ker(F)$.

P5.13 Let $F : \mathbb{R}^5 \to \mathbb{R}^4$ be a linear transformation with $\text{nullity}(F) = 2$. Show that $\text{im}(F)$ is isomorphic to $\mathbb{R}^3$.

P5.14 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be an injective linear transformation with $n \leq m$, and suppose that V is a vector subspace of $\mathbb{R}^n$ with $\dim(V) = k$. Show that $T(V)$ is isomorphic to $\mathbb{R}^k$, where $T(V) = \{T(\vec{v}) : \vec{v} \in V\}$.

Chapter 6

Inverses

6.1. Injective and Surjective Functions

Theorem 5.4 from the previous chapter tells us that we can apply the theory we've developed in the course so far to study linear transformations. In Chapter Exercise 5.2, you showed that linear transformations map vector spaces to vector spaces. In this Chapter, we'll gather tools to detect how the dimension of a vector space may change under a linear function.

We defined the following.

Definition 6.1. Let $f : X \to Y$ be a function for sets X and Y.

(1) f is called ONE-TO-ONE (or INJECTIVE) if the following property holds: for every $y \in Y$, there is *at most* one input $x \in X$ so that $f(x) = y$. We often use the arrow $f : X \hookrightarrow Y$ to indicate when a function is injective.

(2) f is called ONTO (or SURJECTIVE) if the following property holds: for every vector $\vec{b}$ in $\mathbb{R}^n$ there is *at least* one vector $\vec{x}$ in $\mathbb{R}^m$ so that $f(\vec{x}) = \vec{b}$. We often use the arrow $f : \mathbb{R}^m \twoheadrightarrow \mathbb{R}^n$ to indicate when a function is surjective.

(3) f is called BIJECTIVE if it is both one-to-one and onto.

Example 6.2. This may be familiar from your calculus courses for functions between one-dimensional Euclidean space. For example, the function

$$f : \mathbb{R} \to \mathbb{R}, x \mapsto 2x + 1$$

is one-to-one and onto. The function

$$g : \mathbb{R} \to \mathbb{R}, x \mapsto x^2$$

is neither one-to-one nor onto.

Activity 6.1. Determine whether each of the following functions are injective, surjective, or bijective.

(1) $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ where
$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

(2) $T_B : \mathbb{R}^2 \to \mathbb{R}^2$ where
$$B = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

(3) $T_C : \mathbb{R}^3 \to \mathbb{R}^2$ where
$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

(4) $T_D : \mathbb{R}^2 \to \mathbb{R}^3$ where
$$D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Theorem 6.3. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with defining matrix A_F. Then,*

(1) F is injective if and only if every column in $\operatorname{rref}(A_F)$ has a pivot.

(2) F is surjective if and only if every row in $\operatorname{rref}(A_F)$ has a pivot.

Proof. Let $A = A_F$ be the defining matrix of F. Take any $\vec{y} \in \mathbb{R}^m$ and consider the system
$$\left(A \mid \vec{y}\right).$$
By Rouché-Capelli (Theorem 1.26) this system has at most one solution if and only if $\operatorname{rref}(A)$ has a pivot in every column, as needed. Next, by Theorem 4.7 we know that $\operatorname{rank}(A)$ is equal to the number of pivot columns of A. If $\operatorname{rref}(A)$ has a pivot in every row, then since A is an $m \times n$ matrix we have that $\operatorname{rank}(A) = m$ and so $\operatorname{im} F$ if an m-dimensional subspace of $\mathbb{R}^m$, which gives that $\operatorname{im} F = \mathbb{R}^m$ as needed (this can be seen as a consequence of P3.3). Now, if A has a row without a pivot, then $\operatorname{rank}(A) < m$ and so $\operatorname{im}(F)$ is a k-dimensional subspace of $\mathbb{R}^m$ with $k < m$, which means that $\operatorname{im}(F) \neq \mathbb{R}^m$, and so F is not surjective. $\square$

Activity 6.2. Determine whether each of the following functions are injective, surjective, or bijective.

(1) $F : \mathbb{R}^2 \to \mathbb{R}^3$, $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$

(2) $G : \mathbb{R}^3 \to \mathbb{R}^3$, $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x - y \\ y + z \\ x + z \end{pmatrix}$

(3) $H : \mathbb{R}^3 \to \mathbb{R}^3, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x-y \\ y+z \\ z \end{pmatrix}$

6.2. Isomorphisms

In Chapter Exercise 6.1, you'll show the following.

Proposition 6.4. Suppose that $F : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. If F is bijective, then we must have $n = m$.

This proposition tells us that bijective linear functions can only map between "identical" Euclidean spaces. This idea generalizes to bijective linear maps between vector spaces more generally. So far, we've been focusing on linear maps between Euclidean spaces $\mathbb{R}^n \to \mathbb{R}^m$. Note that we can also define a linear transformation between vectors subspaces $V \subseteq \mathbb{R}^n$ and $W \subseteq \mathbb{R}^m$.

Definition 6.5. Let V be a subspace of $\mathbb{R}^n$ and W a subspace of $\mathbb{R}^m$. An ISOMORPHISM between V and W is any linear bijective map $F : V \to W$. If an isomorphism exists between two vector spaces, we say these spaces are ISOMORPHIC, and we write $V \cong W$.

We have the following.

Theorem 6.6. *Let V and W be vector subspaces. Then $V \cong W$ if and only if* $\dim(V) = \dim(W)$.

Proof. Suppose that $F : V \to W$ is an isomorphism, and suppose that V has basis $\{\vec{v}_1, \dots, \vec{v}_d\}$. Note that this implies $\dim(V) = d$. Let's show that the set $\{F(\vec{v}_1), \dots, F(\vec{v}_d)\}$ is a basis for W. To see that this set is linearly independent, suppose that

$$x_1 F(\vec{v}_1) + \cdots + x_d F(\vec{v}_d) = \vec{0}.$$

Since F is linear, this gives

$$F(x_1\vec{v}_1 + \cdots + x_d\vec{v}_d) = \vec{0}.$$

But since F is one-to-one and $F(\vec{0}) = \vec{0}$ we must have

$$x_1\vec{v}_1 + \cdots + x_d\vec{v}_d = \vec{0}.$$

Since the $\vec{v}_i$ are linearly independent, we must have $x_1 = \cdots = x_d = 0$. Hence, $F(\vec{v}_1), \dots, F(\vec{v}_d)$ are linearly independent as well. Next, to see that this set spans W, take any $\vec{w} \in W$. Since F is onto, there exists $\vec{v} \in V$ so that $F(\vec{v}) = \vec{w}$. Since the $\vec{v}_i$ form a basis for V we can write $\vec{v} = c_1\vec{v}_1 + \cdots + c_d\vec{v}_d$ and so

$$\begin{aligned} \vec{w} &= F(c_1\vec{v}_1 + \cdots + c_d\vec{v}_d) \\ &= c_1F(\vec{v}_1) + \cdots + c_dF(\vec{v}_d) \\ &\in \operatorname{Span}(F(\vec{v}_1), \dots, F(\vec{v}_d)), \end{aligned}$$

where the second equality follows by linearity of F. Hence, $W = \operatorname{Span}(F(\vec{v}_1), \dots, F(\vec{v}_d))$ and so the $F(\vec{v}_i)$ form generating set. Hence, $\{F(\vec{v}_1), \dots, F(\vec{v}_d)\}$ is a basis for W, and so $\dim(W) = d$.

Conversely, suppose that $\dim(V) = \dim(W)$. Let
$$\{\vec{v}_1, \ldots, \vec{v}_d\}$$
be a basis for V and
$$\{\vec{w}_1, \ldots, \vec{w}_d\}$$
be a basis for W. Define the map $F : V \to W$ by
$$F(x_1\vec{v}_1 + \cdots + x_d\vec{v}_d) = x_1\vec{w}_1 + \cdots + x_d\vec{w}_d.$$
(that is, $F : \vec{v}_i \mapsto \vec{w}_i$ and we extend linearly). It can be checked by definition that F is an isomorphism. $\square$

Example 6.7. Consider the spaces
$$V = \operatorname{Span}\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right) \text{ and } W = \operatorname{Span}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right).$$
Then $\dim(V) = \dim(W)$ and so $V \cong W$. Furthermore, both V and W are isomorphic to $\mathbb{R}^2$. In this case $V = \mathbb{R}^2$, and while W isn't **equal** to $\mathbb{R}^2$ on the nose (since it contains three-dimensional vectors) it "looks like" $\mathbb{R}^2$ (which is what isomorphisms are meant to capture).

6.3. Matrix Inverses

Recall the following definitions.

Definition 6.8. Let X be a set. Then, the IDENTITY on X is the function
$$\mathrm{id}_X : X \to X$$
defined by $\mathrm{id}_X(x) = x$ for all $x \in X$.

Definition 6.9. Let $f : X \to Y$ be a function. Then, the INVERSE of f (if it exists) is the function $f^{-1} : Y \to X$ so that
$$f^{-1} \circ f = \mathrm{id}_X \text{ and } f \circ f^{-1} = \mathrm{id}_Y .$$
If such a function exists, we say that f is INVERTIBLE.

Example 6.10. If $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = 2x+1$, then $f^{-1} : \mathbb{R} \to \mathbb{R}$ exists and is equal to $f^{-1}(y) = \frac{y-1}{2}$.

We have the following observation.

Proposition 6.11. If $F : \mathbb{R}^n \to \mathbb{R}^m$ is a bijective linear transformation, then it's invertible, and $F^{-1} : \mathbb{R}^m \to \mathbb{R}^n$ is also a linear transformation.

Proof. Suppose that $F : \mathbb{R}^n \to \mathbb{R}^m$ is bijective. Since F is bijective, for every $\vec{y} \in \mathbb{R}^m$ there exists a unique vector $\vec{x} \in \mathbb{R}^n$ so that $F(\vec{x}) = \vec{y}$. So, if we define
$$G : \mathbb{R}^m \to \mathbb{R}^n, \vec{y} \mapsto \vec{x}$$
then G is a well-defined function satisfying $(G \circ F)(\vec{x}) = \vec{x}$ and $(F \circ G)(\vec{y}) = \vec{y}$ for all $\vec{x} \in \mathbb{R}^n$ and $\vec{y} \in \mathbb{R}^m$. Hence, F has inverse G, and so F is invertible.

To see that F^{-1} is linear, take any $\vec{y}, \vec{z} \in \mathbb{R}^m$. Since F is surjective, there exists vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ so that $F(\vec{u}) = \vec{y}$ and $F(\vec{v}) = \vec{z}$. So, we have

$$\begin{aligned} F^{-1}(\vec{y} + \vec{z}) &= F^{-1}(F(\vec{u}) + F(\vec{v})) \\ &= F^{-1}(F(\vec{u} + \vec{v})), \text{ since } F \text{ is linear} \\ &= \vec{u} + \vec{v}, \text{ since } F^{-1} \text{ is the inverse of } F \\ &= F^{-1}(\vec{y}) + F^{-1}(\vec{z}), \end{aligned}$$

where the final equality follows because $F(\vec{u}) = \vec{y} \Rightarrow F^{-1}(\vec{y}) = \vec{u}$ and $F(\vec{v}) = \vec{z} \Rightarrow F^{-1}(\vec{z}) = \vec{v}$. Similarly, for any $c \in \mathbb{R}$ we have

$$\begin{aligned} F^{-1}(c\vec{y}) &= F^{-1}(cF(\vec{u})) \\ &= F^{-1}(F(c\vec{u})) \\ &= c\vec{u} \\ &= cF^{-1}(\vec{y}), \end{aligned}$$

as needed. □

In this section, we derive a method to find the defining matrix of F^{-1} when F is linear. We first define the following.

Definition 6.12. The IDENTITY MATRIX I_n is the defining matrix of the identity transformation $\mathrm{id}_{\mathbb{R}^n} : \mathbb{R}^n \to \mathbb{R}^n, \mathrm{id}_{\mathbb{R}^n}(\vec{x}) = \vec{x}$.

Activity 6.3. Find I_2, I_3 and I_4. Then, give a description for what I_n looks like for a general integer n.

Definition 6.13 (Inverse Matrix, geometric definition)**.** Let A be an $n \times n$ matrix. Then INVERSE OF A, if it exists, is the defining matrix of the inverse transformation T_A^{-1} (which we know exists by Proposition 6.11).

Observe that if B is the defining matrix of T_A^{-1} then we have

$$T_A \circ T_B = \mathrm{id}_{\mathbb{R}^n} \text{ and } T_B \circ T_A = \mathrm{id}_{\mathbb{R}^n} .$$

By definition of matrix products, this gives

$$AB = BA = I_n.$$

This allows us to give an **equivalent definition** of matrix inverses completely algebraically.

Definition 6.14 (Inverse Matrix, algebraic definition)**.** Let A be an $n \times n$ matrix. Then, the INVERSE OF A, if it exists, is the $n \times n$ matrix B satisfying

$$AB = BA = I_n$$

If such a matrix B exists, we say that the matrix A is INVERTIBLE, and we write $B = A^{-1}$.

As we discussed above, observe that $T_{A^{-1}} = T_A^{-1}$.

Activity 6.4. Consider the matrix

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

Verify that

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

Example 6.15. Let's look at how we could *find* the inverse of the matrix A from Activity 6.4. Observe that for A^{-1} to exist, it must satisfy

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

and so A^{-1} must have column vectors $\vec{b}_1, \vec{b}_2, \vec{b}_3$ satisfying the matrix-vector equations

$$A\vec{b}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, A\vec{b}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \text{ and } A\vec{b}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

We can use row reduction to solve the first matrix-vector equation, as below

$$\begin{aligned}
\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right) &\sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_2 \\
&\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_1 \\
&\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right), \text{ subtracting } R_2 \text{ from } R_3
\end{aligned}$$

This gives

$$\vec{b}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Observe that we can use exactly the same row operations to solve for $\vec{b}_2$ and $\vec{b}_3$:

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_2$$
$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_1$$
$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right), \text{ subtracting } R_2 \text{ from } R_3$$

so that $\vec{b}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ and similarly we compute

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 1 & 1 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_2$$
$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 1 & 1 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_1$$
$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right), \text{ subtracting } R_2 \text{ from } R_3$$

so that $\vec{b}_3 = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$, which gives

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

Note that, since we performed the same row operations for every matrix-vector equation, we could performed the same operations as above by instead looking at the augmented matrix $\left(A \mid I_3\right)$, as follows

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_2$$
$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_1$$
$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & -1 & 2 \end{array}\right), \text{ subtracting } R_2 \text{ from } R_3$$

Now, if we let

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

the from our work above we know that $AB = I_3$. We can also check that $BA = I_3$ and so $B = A^{-1}$.

The following lemma tells us that we only need to check for "one-sided" inverses, which saves us some time.

Lemma 6.16. *For $n \times n$ matrices A, B, if $AB = I_n$ then $B = A^{-1}$.*

Proof. Observe that $\mathrm{Nul}(B) \subseteq \mathrm{Nul}(AB)$, since if $B\vec{x} = \vec{0}$ then

$$(AB)\vec{x} = A(B\vec{x}) = A\vec{0} = \vec{0}.$$

So, $\mathrm{nullity}(B) \leq \mathrm{nullity}(AB) = \mathrm{nullity}(I_n) = 0$. Hence, every column of $\mathrm{rref}(B)$ has a pivot. But since B is square this implies that every row of B must also have a pivot. So, $B \sim I_n$, and in Theorem 6.21 we'll show this implies B is invertible. So we have

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

and so $A = B^{-1}$. □

Theorem 6.17. *Let A be an $n \times n$ matrix. If $\left(A \mid I_n\right)$ is row equivalent to $\left(I_n \mid B\right)$ for an $n \times n$ matrix B, then then A is invertible with $A^{-1} = B$.*

Proof. Suppose that $\left(A \mid I_n\right)$ is row equivalent to $\left(I_n \mid B\right)$ for an $n \times n$ matrix B, and and write

$$B = \begin{pmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_n \end{pmatrix}$$

Let $\vec{u}_i$ be the vector in $\mathbb{R}^n$ with 1 in the ith component and 0s everywhere else. From above, we have that $\left(A \mid \vec{u}_i\right)$ is row equivalent to $\left(I_n \mid \vec{b}_i\right)$, and so $\vec{b}_i$ is a solution to the matrix-vector equation $A\vec{x} = \vec{u}_i$. This gives $AB = I_n$. Next, observe that $\left(A \mid I_n\right)$ being row equivalent to $\left(I_n \mid B\right)$ implies that $\left(B \mid I_n\right)$ is row equivalent to $\left(I_n \mid A\right)$ (this is not immediate, you may try convincing yourself of this with some examples). This proof will then be completed with the following lemma. □

Activity 6.5. Use Theorem 6.17 to find the inverse of

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

6.4. Elementary Matrices

In Theorem 6.17 we saw that if an $n \times n$ matrix A is invertible then $A \sim I_n$. Today, we'll show that the converse is also true. To prove this, it will be convenient to introduce a new definition.

Definition 6.18. An $n \times n$ matrix is called ELEMENTARY if it can be obtained by performing exactly one row operation to the identity matrix.

Since we have three elementary row operations, there should be three types of elementary matrices. We have the following

- **Row-switching matrices**: let S_{ij} be the matrix which is obtained by swapping the ith and jth rows of the identity matrix. For example, for 3×3 matrices we have
$$S_{1,3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$
- **Row-multiplying matrix:** Let $M_i(c)$ be the matrix which is obtained by multiplying the ith row of the identity matrix by a constant c. For example, for 3×3 matrices we have
$$M_2(5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$
- **Row-addition matrix:** Let $A_{i,j}(c)$ be the matrix which is obtained by adding c times the jth row to the ith row of the identity matrix. For example, for 3×3 matrices we have
$$A_{1,2}(5) = \begin{pmatrix} 1 & 5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Activity 6.6. Consider the matrix
$$B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$
Calculate the following matrix products $S_{1,3}B$, $M_2(5)B$, and $A_{1,2}(5)B$. What do you notice?

Our observation from the previous activity generalizes. We omit the proof of this result here, since the notation makes this quite messy to write down formally.

Proposition 6.19. Let E be the elementary matrix obtained by performing row operation $(*)$ to I_n, and let B be any $n \times n$ matrix. Then EB is the matrix obtained by performing elementary row operation $(*)$ to B. More precisely, we have

(1) The matrix $S_{i,j}B$ is equal to the matrix obtained by swapping the ith and jth rows of B;

(2) The matrix $M_i(c)B$ is equal to the matrix obtained by multiplying the ith row of B by constant c;

(3) The matrix $A_{i,j}(c)B$ is equal to the matrix obtained by adding c times the jth row of B to the ith row of B.

Activity 6.7. Let's show that every elementary matrix is invertible.

(1) Find the inverse of the 3×3 elementary matrices

$$S_{1,3}, M_2(5), \text{ and } A_{1,2}(5).$$

(2) Find the inverse of the $n \times n$ elementary matrices

$$S_{ij}, M_i(c), \text{ and } A_{i,j}(c),$$

where c is any nonzero real number.

From the previous activity, we have the following.

Proposition 6.20. Every elementary matrix is invertible. Furthermore, the inverse of an elementary matrix is an elementary matrix.

6.5. The Invertible Matrix Theorem

The following result combines different perspectives we've encountered so far this semester.

Theorem 6.21 (Invertible Matrix Theorem)**.** *Let A be an $n \times n$ matrix. Then, the following are equivalent:*

(1) A is invertible;

(2) The matrix-vector equation $A\vec{x} = \vec{b}$ has a unique solution for any $\vec{b} \in \mathbb{R}^n$;

(3) $\operatorname{rref}(A) = I_n$*;*

(4) A is a product of elementary matrices.

We need one quick lemma.

Lemma 6.22. *For $n \times n$ invertible matrices A, B we have*

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

Proof. We have

$$(AB)(B^{-1}A^{-1}) = AI_nA^{-1} = AA^{-1} = I_n$$
$$(B^{-1}A^{-1})(AB) = B^{-1}I_nB = B^{-1}B = I_n$$

so $(AB)^{-1} = B^{-1}A^{-1}$. □

We are now ready to prove the Invertible Matrix Theorem.

Proof. We'll prove $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (1)$.

$(1) \Rightarrow (2)$: Suppose that A is invertible. Observe that

$$A(A^{-1}\vec{b}) = (AA^{-1})\vec{b} = I_n\vec{b} = \vec{b},$$

and so $A^{-1}\vec{b}$ is a solution to the matrix-vector equation $A\vec{x} = \vec{b}$. To see this solution is *unique*, suppose that we have another solution $\vec{y}$. Then,

$$A\vec{y} = \vec{b} \text{ and } A(A^{-1}\vec{b}) = \vec{b}$$

$$\Rightarrow A\vec{y} = A(A^{-1}\vec{b}).$$

Multiplying the above equation on both sides by A^{-1} gives

$$A^{-1}A\vec{y} = A^{-1}A(A^{-1}\vec{b})$$
$$\Rightarrow \vec{y} = A^{-1}\vec{b},$$

showing uniqueness.

(2) $\Rightarrow$ (3): Since the matrix-vector equation $A\vec{x} = \vec{b}$ has a unique solution, by Rouché-Capelli (Theorem 1.26) rref(A) must have a pivot in every column, and so rref(A) has n pivots. But since A is $n \times n$, rref(A) must also have a pivot in every row. The only $n \times n$ matrix with a pivot in every column and row is I_n.

(3) $\Rightarrow$ (4): Since A is row equivalent to I_n, there is a series of elementary row operations which transform A to I_n. This is equivalent to the equality

$$I_n = E_\ell \cdots E_1 A$$

where E_i are elementary matrices. By Proposition 6.20, each of the elementary matrices E_i are invertible, and so by repeated use of Lemma 6.22 we have

$$A = E_1^{-1} \cdots E_\ell^{-1}.$$

Finally, by Proposition 6.20 we know that E_i^{-1} is elementary for every $i = 1, \ldots, \ell$, and so A is a product of elementary matrices, as needed.

(4) $\Rightarrow$ (1): If $A = E_1 E_2 \cdots E_m$ for elementary matrices E_i. then by Lemma 6.22

$$A^{-1} = E_m^{-1} \cdots E_1^{-1},$$

and so A is invertible (since the inverse exists!). □

Exercises

P6.1 Prove Proposition 6.4. Note that you may not use Theorem 6.6.

P6.2 Is it possible for an $m \times n$ matrix A to have an inverse when $m \neq n$? Explain why or why not.

P6.3 Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Use Theorem 6.17 to show that A is invertible if and only if $ad - bc \neq 0$, and in this case we have

$$A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

P6.4 Show that if a matrix A is invertible, then its inverse A^{-1} is unique. That is, if there exist matrices B and C satisfying

$$AB = BA = I_n \text{ and } AC = CA = I_n$$

show that $B = C$.

P6.5 Prove that the following conditions are equivalent. *(Hint: use the Invertible Matrix Theorem, Theorem 6.21)*

(a) A is invertible;
(b) $\mathrm{rref}(A)$ has n pivots;
(c) $\mathrm{Nul}(A) = \{\vec{0}\}$;
(d) $\mathrm{Col}(A) = \mathbb{R}^n$;
(e) The columns of A are linearly independent;
(f) T_A is an isomorphism;
(g) T_A is injective;
(h) T_A is surjective.

Chapter 7

Determinants

In this chapter we'll define a powerful tool called *the determinant*. To understand why this tool is so powerful, we'll build the determinant completely geometrically in $\mathbb{R}^2$ and $\mathbb{R}^3$, and then extend our definition algebraically to higher dimensions.

7.1. Determinants in $\mathbb{R}^2$

Recall (for example, from Activity 5.3 and Theorem 5.4) that a linear transformation is completely determined by where we send the standard basis. That is, if we know where a linear function F sends $\vec{e}_1, \ldots, \vec{e}_n$ (or any other basis for that matter) then we know how the function behaves on all of $\mathbb{R}^n$. Let's use the observation to visualize how a function behaves on its entire domain space. We define the following.

Definition 7.1. The UNIT SQUARE is the subset of $\mathbb{R}^2$ given by

$$S := \{\alpha_1\vec{e}_1 + \alpha_2\vec{e}_2 : 0 \leq \alpha_1, \alpha_2 \leq 1\}.$$

Activity 7.1. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by

$$F(\vec{e}_1) = 3\vec{e}_1 \text{ and } F(\vec{e}_2) = 2\vec{e}_2.$$

(1) Sketch a picture of the unit square S.
(2) Sketch a picture of $F(S) := \{F(\vec{v}) : \vec{v} \in S\}$ as a subset of $\mathbb{R}^2$.
(3) Sketch the image of the "standard coordinate grid" for $\mathbb{R}^2$ under F.

Activity 7.2. Let $G : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by

$$G(\vec{e}_1) = \vec{e}_1 + \vec{e}_2 \text{ and } G(\vec{e}_2) = 2\vec{e}_2.$$

(1) Sketch a picture of $G(S)$ as a subset of $\mathbb{R}^2$.
(2) Sketch the image of the "standard coordinate grid" for $\mathbb{R}^2$ under G.

The previous activities gave examples of the following.

Proposition 7.2. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation. Then $F(S)$ is a parallelogram with sides given by $F(\vec{e}_1)$ and $F(\vec{e}_2)$. Furthermore, the standard coordinate grid for $\mathbb{R}^2$ is transformed into a grid with axes in the $F(\vec{e}_1)$ and $F(\vec{e}_2)$ directions.

Proof. We have

$$\begin{aligned} F(S) &= \{F(\alpha_1\vec{e}_1 + \alpha_2\vec{e}_2) : 0 \le \alpha_1, \alpha_2 \le 1\} \\ &= \{\alpha_1 F(\vec{e}_1) + \alpha_2 F(\vec{e}_2) : 0 \le \alpha_1, \alpha_2 \le 1\} \end{aligned}$$

which gives a parallelogram with sides defined by $F(\vec{e}_1)$ and $F(\vec{e}_2)$ We omit a formal proof of the second claim here, and instead invoke the same geometric reasoning we saw in Activities 7.1 and 7.2. □

Activity 7.3. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation.

(1) Suppose that F is surjective. Use geometric reasoning to argue that the area of $F(S)$ is positive.
(2) Suppose that F is not surjective. Use geometric reasoning to argue that the area of $F(S)$ is equal to zero.
(3) Suppose that the area of $F(S)$ is equal to 6. What will the image of the standard coordinate grid under F look like?
(4) Suppose that the area of $F(S)$ is equal to 1/10. What will the image of the standard coordinate grid under F look like?

The activity above demonstrates that we can gain geometric information about a linear transformation by just considering the area of the image of the unit square. Let's add one more piece of geometric information to this story.

Definition 7.3. An ordered basis $\{\vec{b}_1, \vec{b}_2\}$ for $\mathbb{R}^2$ is called POSITIVELY ORIENTED if we can rotate $\vec{b}_1$ less than 180° counterclockwise to reach $\vec{b}_2$. Otherwise, the basis is called NEGATIVELY ORIENTED.

Activity 7.4. Find the orientation for the following ordered bases for $\mathbb{R}^2$.

(1) $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}$ where

$$\vec{b}_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \text{ and } \vec{b}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

(2) $\mathcal{C} = \{\vec{c}_1, \vec{c}_2\}$ where

$$\vec{c}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \vec{c}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

We are now prepared to define the determinant in $\mathbb{R}^2$.

Definition 7.4. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation. Then, the DETERMINANT of F, denoted by $\det(F)$, is the *oriented area* of $F(S)$. That is, if we let

$a(F(S))$ denote the area of $F(S)$ then

$$\det(F) := \begin{cases} a(F(S)) & \text{if } \{F(\vec{e}_1), F(\vec{e}_2)\} \text{ is positively oriented} \\ -a(F(S)) & \text{if } \{F(\vec{e}_1), F(\vec{e}_2)\} \text{ is negatively oriented} \\ 0 & \text{if } a(F(S)) = 0. \end{cases}$$

If A is a 2×2 matrix, we define the DETERMINANT OF A, denoted by $\det(A)$, to be the determinant of the matrix transformation T_A. That is, $\det(A) := \det(T_A)$.

Activity 7.5. Find the determinant of the matrices

$$A = \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Activity 7.6. In this activity, we develop a method to calculate the determinant of a 2×2 matrix completely algebraically. Let

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

For simplicity we'll assume that the set $\left\{ \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix} \right\}$ is a positively oriented ordered basis with $d \neq 0$ and both vectors in the first quadrant.

(1) Recall that the area of a parallelogram can be computed as the product of its base times its height. Use this observation to calculate the determinant of

$$\begin{pmatrix} e & b \\ 0 & d \end{pmatrix}.$$

(2) Show that

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det \begin{pmatrix} a - \frac{bc}{d} & b \\ 0 & d \end{pmatrix}.$$

(3) Use the previous two parts to conclude that

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

With the appropriate casework, we could use a similar argument to what we need in the previous activity to show this formula holds more generally. We'll allow ourselves to use this result without completing the remaining cases.

Proposition 7.5. $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$ for any $a, b, c, d \in R$.

Activity 7.7. Use geometric reasoning to argue that a 2×2 matrix A is invertible if and only if $\det(A) \neq 0$. Conclude that $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. (Note that this gives an alternate method to prove Chapter Exercise P6.3).

7.2. Determinants in $\mathbb{R}^3$

Note that we can extend all of our definitions from the previous section to three dimensions. We have the following.

Definition 7.6. The UNIT CUBE is the subset of $\mathbb{R}^3$ given by

$$C := \{\alpha_1\vec{e}_1 + \alpha_2\vec{e}_2 + \alpha_3\vec{e}_3 : 0 \leq \alpha_1, \alpha_2, \alpha_3 \leq 1\}.$$

Activity 7.8. Let $H : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation.
(1) Sketch a picture of the unit cube C as a subset of $\mathbb{R}^3$.
(2) What type of geometric object could the set $H(C_3)$ be equal to?

Definition 7.7. An ordered basis $\{\vec{b}_1, \vec{b}_2, \vec{b}_3\}$ for $\mathbb{R}^3$ is called POSITIVELY ORIENTED if it satisfies the RIGHT-HAND RULE. That is, if we point our right hand in the direction of $\vec{b}_1$, and curl our fingers in the direction of $\vec{b}_2$, then our thumb should be pointing in the direction of $\vec{b}_3$" (that is, we can rotate vector $\mathfrak{b}_3$ less than 180° counterclockwise to reach our thumb). Otherwise, the basis is called NEGATIVELY ORIENTED.

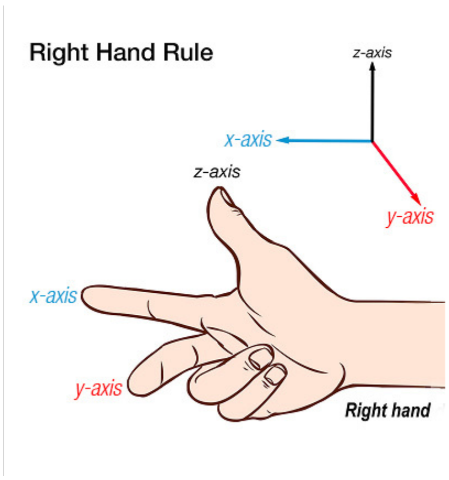


Figure 1. The right-hand rule demonstrating the standard basis is positively oriented (image from PASCO).

Activity 7.9. Find the orientation of the following ordered bases for $\mathbb{R}^3$.
(1) $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \vec{b}_3\}$ where

$$\vec{b}_1 = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}, \vec{b}_2 = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \text{ and } \vec{b}_3 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}.$$

(2) $\mathcal{C} = \{\vec{c}_1, \vec{c}_2, \vec{c}_3\}$ where

$$\vec{c}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \vec{c}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \vec{c}_3 = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}.$$

We can now define the determinant in $\mathbb{R}^3$ similarly to before.

Definition 7.8. Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation. Then, the DETERMINANT of F, denoted by $\det(F)$, is the *oriented volume* of $F(C)$. That is, if we let $v(F(C))$ denote the area of $F(C)$ then

$$\det(F) := \begin{cases} v(F(C)) & \text{if } \{F(\vec{e}_1), F(\vec{e}_2), F(\vec{e}_3)\} \text{ is positively oriented} \\ -v(F(C)) & \text{if } \{F(\vec{e}_1), F(\vec{e}_2), F(\vec{e}_3)\} \text{ is negatively oriented} \\ 0 & \text{if } v(F(C)) = 0. \end{cases}$$

If A is a 3×3 matrix, we define the DETERMINANT OF A, denoted by $\det(A)$, to be the determinant of the matrix transformation T_A. That is, $\det(A) := \det(T_A)$.

To calculate the determinant of 3×3 matrices, we'll need a few observations.

Proposition 7.9. $\det \begin{pmatrix} c & 0 & 0 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} = c \det \begin{pmatrix} a_2 & a_3 \\ b_2 & b_3 \end{pmatrix}$.

Proof. We omit a formal proof here, and note that this follows because the area of a parallelepiped is equal to the area of its base parallelogram times the height. The result follows by considering the base to be the parallelogram in the yz-plane defined by the last two columns, in which case the height is c. □

Proposition 7.10. For any 3×3 matrix $A = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{pmatrix}$ we have

$$\det \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{pmatrix} = -\det \begin{pmatrix} \vec{v}_2 & \vec{v}_1 & \vec{v}_3 \end{pmatrix}.$$

Proof. We omit the details here, and note that the reader can check the needed cases geometrically using the right-hand rule. □

Activity 7.10. Calculate the determinant of the matrices

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

Proposition 7.11. The determinant is row linear. That is

$$\det \begin{pmatrix} \alpha & \beta & \gamma \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} = \det \begin{pmatrix} \alpha & 0 & 0 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} + \det \begin{pmatrix} 0 & \beta & 0 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} + \det \begin{pmatrix} 0 & 0 & \gamma \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix}$$

Proof. We're going to omit this proof, since it's quite messy and not particularly enlightening. If you're interested in this proof, you can checkout the multilinearity property in these notes. We will also aim to revisit this property later on in the course (once we learn about projections in the chapter on orthogonality, we can say something more enlightening here). ☐

Example 7.12. Let's calculate the determinant of the 3×3 matrix

$$A = \begin{pmatrix} a & b & c \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

First, we use Proposition 7.11 to get

$$\det(A) = \det \begin{pmatrix} a & 0 & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \det \begin{pmatrix} 0 & b & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \det \begin{pmatrix} 0 & 0 & c \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

Next, we use Proposition 7.10, which yields

$$\det(A) = \det \begin{pmatrix} a & 0 & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} - \det \begin{pmatrix} b & 0 & 0 \\ 2 & 1 & 3 \\ 5 & 4 & 6 \end{pmatrix} + \det \begin{pmatrix} c & 0 & 0 \\ 3 & 1 & 2 \\ 6 & 4 & 5 \end{pmatrix}.$$

Finally, we can use Proposition 7.9 to get

$$\begin{aligned} \det(A) &= a \det \begin{pmatrix} 2 & 3 \\ 5 & 6 \end{pmatrix} - b \det \begin{pmatrix} 1 & 3 \\ 4 & 6 \end{pmatrix} + c \det \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} \\ &= \boxed{-3a + 6b - 3c}. \end{aligned}$$

Activity 7.11. Use the method in the previous example to calculate $\det(A)$, where

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

Given this calculation, what can you say about the matrix transformation $T_A : \mathbb{R}^3 \to \mathbb{R}^3$?

7.3. Cofactor Expansion and Determinants in $\mathbb{R}^n$

Example 7.12 will help us extend the definition of the determinant to any $n \times n$ matrix. We first need a definition.

Definition 7.13. For an $n \times n$ matrix $A = \left(a_{ij}\right)$, the ij-MINOR of A is defined to be the $(n-1) \times (n-1)$ matrix A_{ij} with the ith row and jth column deleted.

Definition 7.14. Let A be the $n \times n$ matrix with ij-entry equal to a_{ij}. Then we define

$$\det(A) := a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + \cdots + (-1)^{n+1} a_{1n} \det(A_{1n}).$$

Activity 7.12. Find the determinant of the following matrix

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

Remark 7.15. It can be checked that this definition of the determinant agrees with our geometric definitions in $\mathbb{R}^2$ (using Activity **??**) and $\mathbb{R}^3$ (similarly to what was done in Example 7.12). Because of this, we can take an analogous geometric perspective of the determinant in $\mathbb{R}^n$. We define the following.

Definition 7.16. The UNIT n-CUBE C_n is the subset of $\mathbb{R}^n$ defined by

$$C_n = \{\alpha_1 \vec{e}_1 + \cdots \alpha_n \vec{e}_n : 0 \leq \alpha_i \leq n\}.$$

The following generalizes Proposition 7.2.

Proposition 7.17. Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Then, the image of the unit n-cube under F is a parallelepiped in $\mathbb{R}^n$. More precisely,

$$F(C_n) = \{\alpha_1 F(\vec{e}_1) + \cdots + \alpha_n F(\vec{e}_n) : 0 \leq \alpha_1, \ldots, \alpha_n\}$$

is the parallelepiped with sides defined by $F(\vec{e}_1), \ldots, F(\vec{e}_n)$.

We'll see in the next section that the determinant in $\mathbb{R}^n$ behaves quite a lot like the determinants in $\mathbb{R}^3$. Because of this, we *define* the volume of $F(C_n)$ as follows

$$v(F(C_n)) := |\det(F)|.$$

If you go on to take a multivariable calculus course, you'll see that the the determinant can be used to define volumes of subsets of $\mathbb{R}^n$ more generally.

7.4. Properties of the Determinant

Our main goal in this section is to prove the following result, which generalizes what we found geometrically in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Theorem 7.18. *Let A be an $n \times n$ matrix. Then, A is invertible if and only if $\det(A) \neq 0$.*

Since we don't have a concrete notion of geometry in $\mathbb{R}^n$ when $n \geq 4$, we'll need to prove this result algebraically. To do this, we first gather some properties (which, incidentally, will help us simplify future determinant calculations). We first need the following observation.

Lemma 7.19. *Let $A = \begin{pmatrix} a_{ij} \end{pmatrix}$ be an $n \times n$ matrix with ij-entry equal to a_{ij}. Then,*

$$\det(A) = a_{11}\det(A_{11}) - a_{21}\det(A_{21}) + \cdots + (-1)^{n+1} a_{n-1}\det(A_{n1}).$$

That is, we can calculate the determinant by using cofactor expansion along the first column instead of the first row.

To prove this lemma, we would need a technique called *proof by induction*. This method is beyond the scope of this course, so instead we'll convince ourselves of this lemma with a few examples. Students who are interested in a formal proof of this result can see Lemma 4.13 of [**2**].

Activity 7.13. (1) Recall the matrix A from 7.11

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

Use cofactor expansion along the first column of A to verify Lemma 7.19 holds for this example.

(2) Verify Lemma 7.19 holds for an arbitrary 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

We have the following.

Theorem 7.20. *Let A be an $n \times n$ matrix.*

(1) If B is obtained by interchanging two rows of A, then

$$\det(B) = -\det(A);$$

(2) If B is obtained by multiplying one row of A by a constant c, then

$$\det(B) = c\det(A);$$

(3) If B is obtained by replacing a row of A by that row and a scalar multiple of another row of A, then $\det(B) = \det(A)$.

The proof of Theorem 7.20 also needs proof by induction. Let's instead verify this result holds in the 2×2 and 3×3 cases.

Activity 7.14. Use Lemma 7.19 to show the following.

(1) Verify Theorem 7.20 holds for an arbitrary 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

(2) Verify Theorem 7.20 holds for an arbitrary 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

We can now obtain one direction of Theorem 7.18 as a Corollary to Theorem 7.20.

Corollary 7.21. If A is not invertible, then $\det(A) = 0$.

Proof. If A is not invertible, then T_A is not surjective, and so by Theorem 6.3, rref(A) has a row without a pivot. Hence, A is row equivalent to a matrix B where the first row of B is a row of zeros. Hence, by cofactor expansion and Theorem 7.20, we get that $\det(A) = k\det(B) = k \cdot 0 = 0$, where k is some real number constant. □

We need one more result before we're prepared to prove Theorem 7.18.

Lemma 7.22. *Let E be an $n \times n$ elementary matrix, and B be any $n \times n$ matrix. Show that* $\det(EB) = \det(E)\det(B)$.

We leave the proof of this lemma as a Chapter Exercise. We are now prepared to prove our main result.

Proof of Theorem 7.18. In Corollary 7.21 we showed that if $\det(A) \neq 0$ then A is invertible. So, suppose that A is invertible. Then, by the Invertible Matrix Theorem (Theorem 6.5), we can write

$$A = E_1 \cdots E_k$$

where $E_1, \ldots, E_k$ are elementary matrices. Observe that $\det(I_n) = 1$ and so by Proposition 7.20 we get that $\det(E_i) \neq 0$ for $i = 1, \ldots, k$. Hence, by repeated use of Lemma 7.22 we get

$$\det(A) = \det(E_1) \cdots \det(E_k) \neq 0,$$

as needed. □

The following Proposition will help us simplify our determinant calculations.

Proposition 7.23. Let A and B be $n \times n$ matrices. Then,

(1) $\det(A) = \det(A^\top)$ and

(2) $\det(AB) = \det(A)\det(B)$.

Proof. Suppose first that A and B are both invertible. Then, by the Invertible Matrix Theorem (Theorem 6.21) we can write

$$A = E_1 \cdots E_k$$

$$B = E_{k+1} \cdots E_r$$

where $E_1, \ldots, E_r$ are elementary matrices. To prove part (1), we use the properties in Chapter Exercise 7.1. We have $A^\top = E_k^\top \cdots E_1^\top$ and so

$$\begin{aligned}
\det(A^\top) &= \det(E_k^\top) \cdots \det(E_1^\top) \\
&= \det(E_k) \cdots \det(E_1) \\
&= \det(E_1) \cdots \det(E_k) \\
&= \det(E_1 \cdots E_k).
\end{aligned}$$

For part (2), we have $AB = E_1 \cdots E_r$ and so by repeated use of Chapter Exercise 7.1 (c) gives

$$\begin{aligned}\det(AB) &= \det(E_1)\cdots\det(E_r)\\ &= \det(E_1\cdots E_k)\det(E_{k+1}\cdots E_r)\\ &= \det(A)\det(B),\end{aligned}$$

as needed. The remaining case where one of A or B is *not* invertible is covered in Chapter Exercise 7.2. □

Exercises

P7.1 Prove Lemma 7.22. *(Hint: use Proposition 7.20).*

P7.2 This problem will help to finish the proof of Proposition 7.23 by covering properties of the matrix transpose and determinants of elementary matrices.
- (a) Let A and B be $n \times n$ matrices. Show that $(AB)^\top = B^\top A^\top$.
- (b) Show that $\det(E) = \det(E^\top)$ for any elementary matrix E.

P7.3 This problem will help to finish the proof of Proposition 7.23 by discussing the case where one of A or B is not invertible. Note that you may not use Proposition 7.23 to solve this problem. In this problem, we assume that A and B are $n \times n$ matrices.
- (a) Show that if A is not invertible, then $A^\top$ is not invertible. Conclude that if A is not invertible, then $\det(A^\top) = 0$.
- (b) Show that if A or B is not invertible, then AB is not invertible. Conclude that if one of A or B is not invertible, we have $\det(AB) = \det(A)\det(B)$.

P7.4 Let A be an invertible matrix. Show that $\det(A^{-1}) = \dfrac{1}{\det(A)}$.

P7.5 True or False: For any $n\times n$ matrix A, $\det(-A) = -\det(A)$. **If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample**.

P7.6 True or False: For any two $n \times n$ matrices, $\det(A + B) = \det(A) + \det(B)$. **If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample**.

P7.7 True or False: if A and B are square matrices and AB is invertible, then both A and B are invertible. **If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample**.

P7.8 Show that if A and B are $n \times n$ matrices with $\text{rank}(A) = \text{rank}(B) = n$ then $\text{rank}(AB) = n$.

P7.9 An $n \times n$ matrix A is called SKEW-SYMMETRIC if $A = -A^\top$. Show that when n is odd, any $n \times n$ skew-symmetric matrix is not invertible.

P7.10 An $n \times n$ matrix A is called NILPOTENT if A^m is equal to the zero matrix for some positive integer m. Show that nilpotent matrices are not invertible.

Eigenvalues and Eigenvectors

8.1. Definitions

In this Chapter, we'll investigate the "stretch factors" of a linear transformation, which will help us gain a further geometric understanding of how a linear function transforms a vector space. We have the following definition.

Definition 8.1. Let A be an $n \times n$ matrix. A non-zero vector $\vec{x}$ is an EIGENVECTOR of A if there is a real number scalar λ such that $A\vec{x} = \lambda\vec{x}$. The scalar λ is called an EIGENVALUE of A.

Geometrically, this means that when we apply the matrix transformation T_A to an eigenvector $\vec{x}$, this is the same thing as stretching the vector $\vec{x}$ by the eigenvalue λ, as visualized in this 3Blue1Brown video.

Activity 8.1. For each of the following matrix-vector pairs, determine whether $\vec{x}$ is an eigenvector of the matrix A. If it is find the corresponding eigenvalue λ.

(1) $A = \begin{pmatrix} 3 & 2 \\ 3 & 8 \end{pmatrix}$ and $\vec{x} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$.

(2) $B = \begin{pmatrix} 3 & 2 \\ 3 & 8 \end{pmatrix}$ and $\vec{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

(3) $C = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ and $\vec{x} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

Let's develop a method to calculate the eigenvalues and eigenvectors of a given matrix. We have the following.

Proposition 8.2. For an $n \times n$ matrix A, the set of eigenvectors of A corresponding to an eigenvalue λ is equal to the nonzero vectors in $\mathrm{Nul}(A - \lambda I_n)$.

Proof. Observe that any $\vec{x} \in \mathrm{Nul}(A - \lambda I_n)$ if and only if

$$\begin{aligned} &(A - \lambda I_n)\vec{x} = \vec{0} \\ \Leftrightarrow\ & A\vec{x} - \lambda\vec{x} = \vec{0} \\ \Leftrightarrow\ & A\vec{x} = \lambda\vec{x}. \end{aligned}$$

□

Definition 8.3. We call the space $\mathrm{Nul}(A - \lambda I_n)$ the λ-EIGENSPACE of A, and use the notation $E_\lambda := \mathrm{Nul}(A - \lambda I_n)$. By Proposition 8.2 the nonzero vectors in E_λ is equal to the set of all eigenvectors with corresponding eigenvalue λ. Geometrically, E_λ is the set of vectors $\vec{x} \in \mathbb{R}^n$ so that the matrix transformation T_A stretches $\vec{x}$ by a factor of λ.

Activity 8.2. Find the 2-eigenspace of $A = \begin{pmatrix} 3 & 2 \\ 3 & 8 \end{pmatrix}$.

Since we know how to find bases for null spaces, what's left is to find a method to calculate the eigenvalues of a matrix. We have the following.

Proposition 8.4. A real number λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.

Proof. Observe that $\vec{x}$ is an eigenvalue of A if and only if it's a nonzero solution to the matrix-vector equation $A\vec{x} = \lambda\vec{x}$ which can be rewritten as

$$(A - \lambda I_n)\vec{x} = \vec{0}.$$

By the Invertible Matrix Theorem (Theorem 6.21), the matrix-vector equation above having a nontrivial solution is equivalent to the matrix $A - \lambda I_n$ not being invertible. So, the result follows by Theorem 7.18. □

Definition 8.5. For an eigenvalue λ of a matrix A, the GEOMETRIC MULTIPLICITY of λ is defined to be the dimension of the λ-eigenspace E_λ.

8.2. The Characteristic Polynomial

We define the following.

Definition 8.6. For an $n \times n$ matrix A,

$$\chi_A(x) = \det(A - xI_n).$$

is called the CHARACTERISTIC POLYNOMIAL of A.

Activity 8.3. Find the characteristic polynomial of the following matrices. Then, use Proposition 8.4 to find the eigenvalues of each matrix.

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}, C = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Proposition 8.4 and 8.2 give us the following strategy for finding eigenvectors of a matrix A:

(1) Find all eigenvalues of A by solving the polynomial equation $\chi_A(x) = 0$.
(2) For each eigenvalue λ, calculate the λ-Eigenspace $E_\lambda = \text{Nul}(A - \lambda I_n)$
(3) The set of all eigenvalues is the union of the λ-Eigenspaces in part (2).

Given the discussion above, the following observations will help us gain a better geometric understanding of our eigen-story.

Proposition 8.7. For any $n \times n$ matrix A, the characteristic polynomial $\chi_A(x)$ is a polynomial of degree n.

The formal proof of Proposition 8.7 would use an inductive argument, along with the cofactor expansion formula for the determinant. Instead of worrying about understanding this proof formally, note that:

(1) If we look at the cofactor formula for the determinant, we see that the only operations happening are addition and multiplication, and so we end up with some algebraic expression made up of sums and products of real numbers and our unknown x, which precisely defines a polynomial.
(2) If an $n \times n$ matrix A has diagonal entries $d_1, d_2, \ldots, d_n$, then the highest degree term coming out of the cofactor exapansion will be $(d_1 - x)(d_2 - x)\cdots(d_n - x)$ (convince yourself of this in the 3×3 case). So, the degree of $\chi_A(x)$ will be at most n, and in fact in can be argued that the degree is equal to n (by noting that the remaining summands each have degree strictly smaller than n).

We have the following.

Theorem 8.8 (The Fundamental Theorem of Algebra)**.** *Let $f(x)$ be a polynomial of degree n with coefficients in $\mathbb{R}$. Then, we can write*

$$f(x) = (x - \alpha_1)^{m_1} \cdots (x - \alpha_k)^{m_k} \tag{8.1}$$

where $\alpha_1, \ldots, \alpha_k$ are distinct complex numbers, and $m_i \geq 1$ are integers satisfying $m_1 + \cdots + m_k = n$.

Definition 8.9. With the notation of Theorem 8.8,

(1) We call Equation (8.1) the FACTORIZATION of f;
(2) The complex numbers $\alpha_1, \ldots, \alpha_k$ are called the ROOTS of f;
(3) For each $i \in \{1, \ldots, k\}$, the integer m_i is called the ALGEBRAIC MULTIPLICITY of α_i.

The proof of Theorem 8.8 is beyond the scope of this course, so instead let's look at some examples.

Activity 8.4. Find the factorization and roots of the following polynomials. For each root, find its algebraic multiplicity.
(1) $f(x) = x^2 - 2x - 3$
(2) $g(x) = x^4 - 4x^2 + 4$
(3) $h(x) = x^2 + 1$

By Theorem 8.4, the *real* roots of $\chi_A(x)$ are precisely the eigenvalues of A. So, it makes sense for us to find their algebraic multiplicities. Furthermore, Theorem 8.8 tells us that a polynomial of degree n can have *at most* n distinct roots, and so again by Theorem 8.4, an $n \times n$ matrix can have *at most* n linearly independent eigenvectors. In Chapter 10 we'll explore this idea further. For now, let's practice some calculations.

Activity 8.5. Let
$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$
Find the eigenvalues of A. For each eigenvalue, find its geometric and algebraic multiplicity.

Remark 8.10. In this course, we will always insist that our eigenvalues are *real*, however a lot can be said when we encounter complex roots of $\chi_A(x)$. We'll explore this in the 2×2 case in the Chapter Exercises, but we won't have time for the rest of the story this semester.

Let's look into our example from the previous activity more.

Example 8.11. Consider the matrix
$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$
from Activity **??**. Let's use some eigentheory to understand the matrix transformation T_A. In the previous activity, we found that A has eigenvalues $\lambda = 1$ and $\lambda = 2$. We then computed our eigenspaces
$$E_1 = \text{Nul}(A - I_3) = \text{Span}\left(\vec{v}_1, \vec{v}_2\right),$$
where
$$\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$
and $E_2 = \text{Nul}(A - 2 \cdot I_3) = \text{Span}\left(\vec{v}_3\right)$, where
$$\vec{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$
Observe that $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a basis for $\mathbb{R}^3$. For any $\vec{v} \in \mathbb{R}^3$ write
$$\vec{v} = x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3.$$
Using linearity of the matrix transformation T_A gives
$$\begin{aligned} T_A(\vec{x}) &= x_1 T_A(\vec{v}_1) + x_2 T_A(\vec{v}_2) + x_3 T_A(\vec{v}_3) \\ &= (1)(x_1\vec{v}_1 + x_2\vec{v}_2) + 2(x_3\vec{v}_3). \end{aligned}$$
where the second equality follows by recalling that $\vec{v}_1, \vec{v}_2 \in E_1$ and $\vec{v}_3 \in E_2$. This tells us that T_A is the transformation that leaves vectors in the $\vec{v}_1, \vec{v}_2$-directions

fixed, and scales vectors in the $\vec{v}_3$-direction by 2. This gives us a complete picture of how T_A transforms $\mathbb{R}^3$. We'll explore further in Chapter 10, but first we need to develop a little more theory.

Exercises

P8.1 Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation which rotates every vector in $\mathbb{R}^2$ counterclockwise by an angle of θ.

(a) Show that $A_F = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

(b) Show that A_F does not have any (real) eigenvalues unless θ is an integer multiple of $180°$.

P8.2 Show that for any $n \times n$ matrix A we have $\chi_A = \chi_{A^\top}$. Conclude that A and $A^\top$ have the same eigenvalues.

P8.3 True or False: for any $n \times n$ matrix A, A and $A^\top$ have the same eigenvectors. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P8.4 Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_k$. Show that if λ_i has algebraic multiplicity m_i for $i = 1, \ldots, k$, then $\det(A) = \lambda_1^{m_1} \cdots \lambda_k^{m_k}$. That is, the determinant of a matrix is equal to the product of its eigenvalues with multiplicity. *Hint: observe that we can write*

$$\chi_A(x) = \pm(x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}.$$

P8.5 Show that a matrix A is invertible if and only if 0 is not an eigenvalue of A.

P8.6 An $n \times n$ matrix is called IDEMPOTENT if $A^2 = A$.

(a) Show that an idempotent matrix A is invertible if and only if $A = I_n$.

(b) Show that if A is an idempotent matrix that's not equal to the identity then A has eigenvalues 0 and 1.

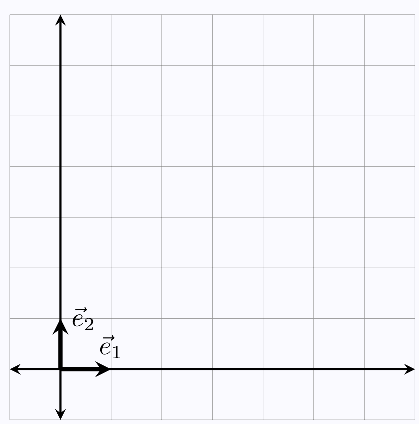
Chapter 9

Coordinate Systems and Change of Basis

9.1. Coordinate Systems

While the number of elements in a basis is fixed, we've seen that there are many (in fact, infinitely many) choices for a basis of a given vector space. In this section, we look at how our choice of basis impacts our geometric understanding of a given vector space.

Activity 9.1. The standard coordinate grid for $\mathbb{R}^2$ is drawn below. Explain how we can use the standard basis $\mathcal{E} = \{\vec{e}_1, \vec{e}_2\}$ to draw the grid lines below.

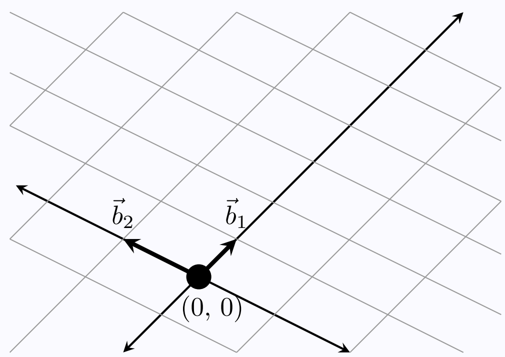


Then, draw the vectors $\vec{u} = 4\vec{e}_1 + \vec{e}_2$ and $\vec{v} = -\vec{e}_1 + 5\vec{e}_2$ on the coordinate grid above.

Activity 9.2. Let $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}$

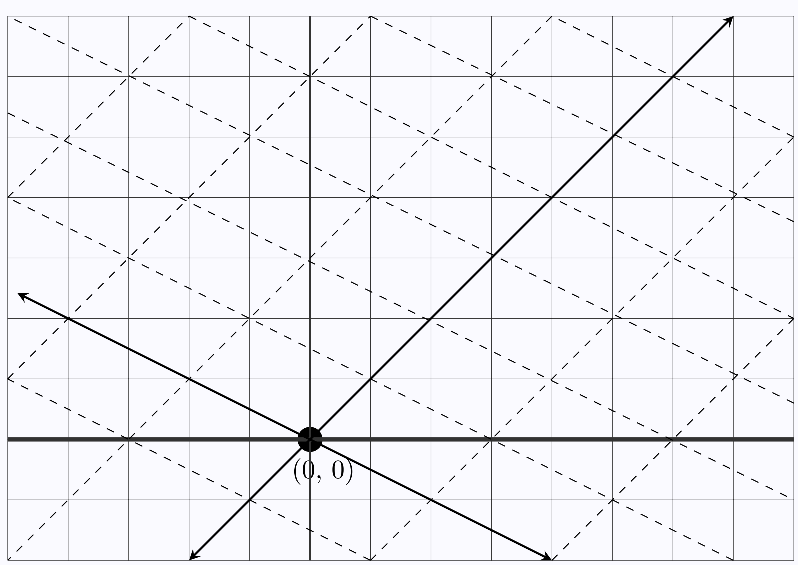
$$\vec{b}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \vec{b}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Observe that $\mathcal{B}$ is a basis for $\mathbb{R}^2$. The image below shows how we can use $\mathcal{B}$ to create a "nonstandard" coordinate grid for $\mathbb{R}^2$. Explain how the vectors $\vec{b}_1, \vec{b}_2$ can be used to draw the grid lines below.



Then, draw the vectors $\vec{w} = 2\vec{b}_1 - \vec{b}_2$ and $\vec{z} = 3\vec{b}_1 + 2\vec{b}_2$ on the coordinate grid above.

Activity 9.3. The following graph includes the standard coordinate grid defined by the standard basis $\mathcal{E}$ (drawn with solid lines) and the "nonstandard" coordinate grid defined by the basis $\mathcal{B}$ (drawn with dashed lines) from the previous activities. Draw the vectors $\vec{u}, \vec{v}, \vec{w}$ and $\vec{z}$ from the previous activities on the graph below. What do you notice?



Activity 9.4. Consider the vector $\vec{v} = 3\vec{e}_1 + 6\vec{e}_2$. Use the graph from the previous activity to find real numbers x_1, x_2 so that $\vec{v} = x_1\vec{b}_1 + x_2\vec{b}_2$.

In the activities above, we observed that different bases give different coordinate grids for $\mathbb{R}^2$, and that coordinate grids give instructions for the location of a vector. Let's introduce some terminology to generalize these observations to $\mathbb{R}^n$.

Definition 9.1. Let $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be an ordered basis for a vector space V. Recall that every vector $\vec{x}$ in V can be written in the form

$$\vec{x} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n.$$

The $\mathcal{B}$-COORDINATES of $\vec{x}$ is the vector in $\mathbb{R}^n$ given by

$$[\vec{x}]_{\mathcal{B}} := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Activity 9.5. Let $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \vec{b}_3\}$ be the ordered basis for $\mathbb{R}^3$ where

$$\vec{b}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \vec{b}_2 = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}, \text{ and } \vec{b}_3 = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}.$$

Find the $\mathcal{B}$-coordinates of the vector $\vec{v} = \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix}$.

Note that, in Activity 9.5 there was precisely one way to write $\vec{v}$ as a linear combination of $\vec{b}_1$, $\vec{b}_2$ and $\vec{b}_3$ since the vector equation represented by

$$\left(\begin{array}{cc|c} \vec{b}_1 & \vec{b}_2 & \vec{v} \end{array}\right)$$

had exactly one solution. This turns out to be true in general (and is what allows us to define **the** coordinates of a vector with respect to a basis). We have the following.

Theorem 9.2. *Let V be a vector subspace of $\mathbb{R}^n$ and $\mathcal{B}$ a basis for V. Then, every vector in V has a unique representation in terms of the basis $\mathcal{B}$. That is, if $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_m\}$, then for every $\vec{v} \in V$ there are unique real numbers $x_1, \ldots, x_m$ so that*

$$\vec{v} = x_1\vec{b}_1 + x_2\vec{b}_2 + \cdots + x_m\vec{b}_m$$

Proof. Let $\vec{v} \in V$. Since $\{\vec{b}_1, \ldots, \vec{b}_m\}$ is a basis for V, this set is linearly independent, and so we know that reduced row echelon form of the matrix

$$\begin{pmatrix} \vec{b}_1 & \cdots & \vec{b}_m \end{pmatrix}$$

has a pivot in every column. Furthermore, since $V = \mathrm{Span}(\vec{b}_1, \ldots, \vec{b}_m)$ we know that the system

$$\left(\vec{b}_1 \quad \cdots \quad \vec{b}_m \;\middle|\; \vec{v}\right)$$

must be consistent. Hence, the reduced row echelon form of the matrix above has a pivot in every column except for the last column, and so by Theorem 1.26 there is exactly one solution to the vector equation

$$\vec{v} = x_1\vec{b}_1 + x_2\vec{b}_2 + \cdots + x_m\vec{b}_m$$

as needed. □

We give an alternate proof below to see how we could instead use our algebraic definition of linear independence.

Proof of Theorem 9.2 (version 2). Let $\vec{v} \in V$ and suppose that we can write

$$\vec{v} = x_1\vec{b}_1 + \cdots + x_m\vec{b}_m$$

and

$$\vec{v} = y_1\vec{b}_1 + \cdots + y_m\vec{b}_m.$$

Subtracting these two equations gives

$$\vec{0} = (x_1 - y_1)\vec{v}_1 + \cdots + (x_m - y_m)\vec{v}_m.$$

Since our vectors $\vec{b}_1, \ldots, \vec{b}_m$ are linearly independent, we must have

$$x_i - y_i = 0 \Rightarrow x_i = y_i$$

for all i. So, our two representations of $\vec{v}$ as a linear combination of the vectors $\vec{v}_1, \ldots, \vec{v}_m$ are the same. □

Remark 9.3. Observe that if $\mathcal{E} = \{\vec{e}_1, \ldots, \vec{e}_n\}$ is the standard basis for $\mathbb{R}^n$, then

$$[x_1\vec{e}_1 + \cdots + x_n\vec{e}_n]_{\mathcal{E}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

That is, when we talk about the coordinates of a vector without referencing any specific basis, we really mean the coordinates of that vector *with respect to the standard basis.*

9.2. Change of Basis Matrices

Observe that we can think of the translation between different coordinate systems as a linear transformation. More precisely, if $\mathcal{C} = \{\vec{c}_1, \ldots, \vec{c}_n\}$ and $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_n\}$ are ordered bases for a vector space V, we can define a linear transformation $F : V \to V$ by mapping $\vec{c}_i \mapsto \vec{b}_i$ and "extending linearly"; that is

$$F(x_1\vec{c}_1 + \cdots x_n\vec{c}_n) = x_1\vec{b}_1 + \cdots x_n\vec{b}_n.$$

Observe that F transforms the coordinate grid defined by $\mathcal{C}$ into the coordinate grid defined by $\mathcal{B}$. In this section, we find the defining matrix of such a transformation.

Definition 9.4. Let $\mathcal{C}$ and $\mathcal{B}$ be bases for a vector space V. Then, the CHANGE OF BASIS matrix $M_{\mathcal{C}\leftarrow\mathcal{B}}$ is the matrix satisfying

$$M_{\mathcal{C}\leftarrow\mathcal{B}}[\vec{x}]_{\mathcal{B}} = [\vec{x}]_{\mathcal{C}}$$

for every vector $\vec{x}$ in V.

Activity 9.6. Consider the basis $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}$ for $\mathbb{R}^2$ where

$$\vec{b}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \text{ and } \vec{b}_2 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}.$$

(1) Find $M_{\mathcal{B}\leftarrow\mathcal{E}}$ and $M_{\mathcal{E}\leftarrow\mathcal{B}}$.
(2) Show that $M_{\mathcal{E}\leftarrow\mathcal{B}}^{-1} = M_{\mathcal{B}\leftarrow\mathcal{E}}$.

Our method from the previous activity generalizes. We have the following.

Theorem 9.5. *Suppose that $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_n\}$ is a basis for $\mathbb{R}^n$. Then, for any vector $\vec{x} \in \mathbb{R}^n$ we have*

$$M_{\mathcal{E}\leftarrow\mathcal{B}} = \begin{pmatrix} [\vec{b}_1]_{\mathcal{E}} & \cdots & [\vec{b}_n]_{\mathcal{E}} \end{pmatrix}.$$

Furthermore, $M_{\mathcal{E}\leftarrow\mathcal{B}}$ is invertible and we have $M_{\mathcal{E}\leftarrow\mathcal{B}}^{-1} = M_{\mathcal{B}\leftarrow\mathcal{E}}$.

Proof. This proof will generalize what we did in the previous examples. Suppose that $[\vec{x}]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Then we have $\vec{x} = x_1\vec{b}_1 + \cdots + x_n\vec{b}_n$. Writing this equation in standard coordinates gives

$$\begin{aligned}
[\vec{x}]_{\mathcal{E}} &= \begin{pmatrix} [\vec{b}_1]_{\mathcal{E}} & \cdots & [\vec{b}_n]_{\mathcal{E}} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\
&= \begin{pmatrix} [\vec{b}_1]_{\mathcal{E}} & \cdots & [\vec{b}_n]_{\mathcal{E}} \end{pmatrix} [\vec{x}]_{\mathcal{B}},
\end{aligned}$$

and so $M_{\mathcal{E}\leftarrow\mathcal{B}} = \begin{pmatrix} [\vec{b}_1]_{\mathcal{E}} & \cdots & [\vec{b}_n]_{\mathcal{E}} \end{pmatrix}$, as needed. Next, since $\mathcal{B}$ is linearly independent, by the Invertible Matrix Theorem (Theorem 6.5), the matrix $M_{\mathcal{E}\leftarrow\mathcal{B}}$ is invertible, and so we have

$$[\vec{x}]_{\mathcal{E}} = M_{\mathcal{E}\leftarrow\mathcal{B}}[\vec{x}]_{\mathcal{B}} \Rightarrow [\vec{x}]_{\mathcal{B}} = M_{\mathcal{E}\leftarrow\mathcal{B}}^{-1}[\vec{x}]_{\mathcal{E}}.$$

Hence, $M_{\mathcal{E}\leftarrow\mathcal{B}}^{-1} = M_{\mathcal{B}\leftarrow\mathcal{E}}$. □

This result tells us how to change between the *standard* basis for $\mathbb{R}^n$ and a non-standard basis. To generalize this idea to changing between two arbitrary bases, we need the following lemmas, whose proofs will be left to a Chapter Exercise.

Lemma 9.6. *Let $\mathcal{C}$ be a basis for a vector space V. Then, for any $\vec{x}, \vec{y} \in V$ and scalar $k \in \mathcal{R}$ we have*

$$[\vec{x} + \vec{y}]_{\mathcal{C}} = [\vec{x}]_{\mathcal{C}} + [\vec{y}]_{\mathcal{C}} \textit{ and } [k\vec{x}]_{\mathcal{C}} = k[\vec{x}]_{\mathcal{C}}.$$

Lemma 9.7. *Let V be a vector space with basis $\mathcal{B} = \{\vec{b}_1, \dots, \vec{b}_n\}$. Then, for any basis $\mathcal{C}$ of V, the set $\{[\vec{b}_1]_{\mathcal{C}}, \dots, [\vec{b}_n]_{\mathcal{C}}\}$ is linearly independent.*

We are now prepared to show the following.

Theorem 9.8. *Let V be a vector space with basis $\mathcal{C} = \{\vec{c}_1, \dots, \vec{c}_n\}$. Then, a subset $\mathcal{B} = \{\vec{b}_1, \dots, \vec{b}_n\}$ of V is a basis for V if and only if $\begin{pmatrix}[\vec{b}_1]_{\mathcal{C}} & \cdots & [\vec{b}_n]_{\mathcal{C}}\end{pmatrix}$ is invertible. In this case, we have*

$$M_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{pmatrix}[\vec{b}_1]_{\mathcal{C}} & \cdots & [\vec{b}_n]_{\mathcal{C}}\end{pmatrix}.$$

and furthermore $M_{\mathcal{C}\leftarrow\mathcal{B}}^{-1} = M_{\mathcal{B}\leftarrow\mathcal{C}}$.

Proof. Suppose first that $\mathcal{B}$ is a basis. Then, for any $\vec{x}$ in V, let $[\vec{x}]_{\mathcal{B}} = \begin{pmatrix}x_1\\ \vdots \\ x_n\end{pmatrix}$. Then we have $\vec{x} = x_1\vec{b}_1 + \cdots + x_n\vec{b}_n$. Writing this equation in $\mathcal{C}$-coordinates gives

$$[\vec{x}]_{\mathcal{C}} = [x_1\vec{b}_1 + \cdots + x_n\vec{b}_n]_{\mathcal{C}}.$$

By Lemma 9.6 this gives

$$\begin{aligned}[\vec{x}]_{\mathcal{C}} &= x_1[\vec{b}_1]_{\mathcal{C}} + \cdots + x_n[\vec{b}_n]_{\mathcal{C}}\\ &= \begin{pmatrix}[\vec{b}_1]_{\mathcal{C}} & \cdots & [\vec{b}_n]_{\mathcal{C}}\end{pmatrix}\begin{pmatrix}x_1\\ \vdots \\ x_n\end{pmatrix}\\ &= \begin{pmatrix}[\vec{b}_1]_{\mathcal{C}} & \cdots & [\vec{b}_n]_{\mathcal{C}}\end{pmatrix}[\vec{x}]_{\mathcal{B}},\end{aligned}$$

and so $M_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{pmatrix}[\vec{b}_1]_{\mathcal{C}} & \cdots & [\vec{b}_n]_{\mathcal{C}}\end{pmatrix}$. Now, by Lemma 9.7 we know that the columns of $M_{\mathcal{C}\leftarrow\mathcal{B}}$ are linearly independent, and so by the Invertible Matrix Theorem (Theorem 6.5) $M_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible. So,

$$[\vec{x}]_{\mathcal{C}} = M_{\mathcal{C}\leftarrow\mathcal{B}}[\vec{x}]_{\mathcal{B}} \Rightarrow [\vec{x}]_{\mathcal{B}} = M_{\mathcal{C}\leftarrow\mathcal{B}}^{-1}[\vec{x}]_{\mathcal{C}}.$$

Hence, we have $M_{\mathcal{C}\leftarrow\mathcal{B}}^{-1} = M_{\mathcal{B}\leftarrow\mathcal{C}}$.

Conversely, suppose that the matrix $\begin{pmatrix}[\vec{b}_1]_{\mathcal{C}} & \cdots & [\vec{b}_n]_{\mathcal{C}}\end{pmatrix}$ is invertible. To see that $\mathcal{B}$ is linearly independent, consider the vector equation

$$\alpha_1\vec{b}_1 + \cdots + \alpha_n\vec{b}_n = \vec{0}.$$

Writing this equation in $\mathcal{C}$-coordinates and applying Lemma 9.6 gives

$$\alpha_1[\vec{b}_1]_{\mathcal{C}} + \cdots + \alpha_n[\vec{b}_n]_{\mathcal{C}} = \begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix}.$$

Since $\begin{pmatrix}[\vec{b}_1]_{\mathcal{C}} & \cdots & [\vec{b}_n]_{\mathcal{C}}\end{pmatrix}$ is invertible, its columns must be linearly independent, and so $\alpha_1 = \cdots = \alpha_n = 0$, as needed. Finally, if $\mathcal{B}$ did not span V, then by Chapter Exercise P3.8, we would have a basis for V of dimension larger than n. This contradicts the fact that V is n-dimensional, and so $\mathcal{B}$ is a basis as needed. $\square$

Activity 9.7. Let V be the plane in $\mathbb{R}^3$ spanned by

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

(1) Show that $\mathcal{B}$ is also a basis for V, where

$$\mathcal{B} = \left\{ \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix} \right\}.$$

(2) Find the change of basis matrices $M_{\mathcal{C}\leftarrow\mathcal{B}}$ and $M_{\mathcal{C}\leftarrow\mathcal{B}}$.
(3) Find a new basis $\mathcal{D}$ for V that's not equal to $\mathcal{B}$ or $\mathcal{C}$.

Exercises

P9.1 True or False: If $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_m\}$ is a generating set for a vector space V, then every vector $\vec{v}$ in V has a *unique* representation in terms of $\mathcal{B}$ (that is, there are unique real numbers $x_1, \ldots, x_m$ so that $\vec{v} = x_1\vec{b}_1 + \cdots + x_m\vec{b}_m$). If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P9.2 Prove Lemma 9.6.

P9.3 Prove Lemma 9.7.

P9.4 Let V be a vector subspace of $\mathbb{R}^5$ of dimension 3, and let $\mathcal{B}$ and $\mathcal{C}$ be bases for V. What is the size of the change of basis matrix $M_{\mathcal{C}\leftarrow\mathcal{B}}$? Justify your answer.

P9.5 Let $\mathcal{B}, \mathcal{C}$ and $\mathcal{D}$ be bases for a vector space V. Show that

$$M_{\mathcal{C}\leftarrow\mathcal{B}} M_{\mathcal{B}\leftarrow\mathcal{D}} = M_{\mathcal{C}\leftarrow\mathcal{D}}.$$

Chapter 10

Similarity and Diagonalization

10.1. Defining Matrices

Recall that for any linear transformation $F : \mathbb{R}^n \to \mathbb{R}^m$ there's a matrix A so that $F(\vec{x}) = A\vec{x}$. In Section 5.2, we called this the *defining matrix* of F, and used the notation A_F. Furthermore, we found that

$$A_F = \begin{pmatrix} F(\vec{e}_1) & \cdots & F(\vec{e}_n) \end{pmatrix}.$$

Note that this story was told in terms of the *standard basis*. That is, $[F(\vec{x})]_{\mathcal{E}} = A_F[\vec{x}]_{\mathcal{E}}$, where $A_F = \begin{pmatrix} [F(\vec{e}_1)]_{\mathcal{E}} & \cdots & [F(\vec{e}_n)]_{\mathcal{E}} \end{pmatrix}$. Let's look at what happens if we instead consider our transformation with respect to a basis other than the standard one.

Theorem 10.1. *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and $\mathcal{B}$ be any basis for $\mathbb{R}^n$. Then, there exists a unique $n \times n$ matrix A so that $[F(\vec{x})]_{\mathcal{B}} = A[\vec{x}]_{\mathcal{B}}$. Furthermore, we have $A = \begin{pmatrix} [F(\vec{b}_1)]_{\mathcal{B}} & \cdots & [F(\vec{b}_n)]_{\mathcal{B}} \end{pmatrix}$.*

Proof. This proof will follow similarly to Theorem 5.4. Take any $\vec{x} \in \mathbb{R}^n$ and write $\vec{x} = x_1\vec{b}_1 + \cdots + x_n\vec{b}_n$. Then,

$$\begin{aligned}
[F(\vec{x})]_{\mathcal{B}} &= [F(x_1\vec{b}_1 + \cdots + x_n\vec{b}_n)]_{\mathcal{B}} \\
&= [x_1F(\vec{b}_1) + \cdots + x_nF(\vec{b}_n)]_{\mathcal{B}} && \text{since } F \text{ is linear} \\
&= x_1[F(\vec{b}_1)]_{\mathcal{B}} + \cdots + x_n[F(\vec{b}_n)]_{\mathcal{B}} && \text{by Lemma 9.6} \\
&= \begin{pmatrix} [F(\vec{b}_1)]_{\mathcal{B}} & \cdots & [F(\vec{b}_1)]_{\mathcal{B}} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\
&= \begin{pmatrix} [F(\vec{b}_1)]_{\mathcal{B}} & \cdots & [F(\vec{b}_1)]_{\mathcal{B}} \end{pmatrix} [\vec{x}]_{\mathcal{B}},
\end{aligned}$$

as needed. ☐

Definition 10.2. Let $F:\mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation, and $\mathcal{B}$ be any basis for $\mathbb{R}^n$. Then, the DEFINING MATRIX OF F WITH RESPECT TO THE BASIS $\mathcal{B}$ is the matrix A so that

$$[F(\vec{x})]_{\mathcal{B}} = A[\vec{x}]_{\mathcal{B}}.$$

We use the notation $A = A_{F,\mathcal{B}}$. By Theorem 10.1 we have

$$A_{F,\mathcal{B}} = \begin{pmatrix} [F(\vec{b}_1)]_{\mathcal{B}} & \cdots & [F(\vec{b}_n)]_{\mathcal{B}} \end{pmatrix}.$$

Activity 10.1. Let $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}$ be the basis with

$$\vec{b}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \vec{b}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

(1) Let $F:\mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 2x+y \\ x-y \end{pmatrix}.$$

Find the defining matrices A_F and $A_{F,\mathcal{B}}$.

(2) Let $G:\mathbb{R}^2 \to \mathbb{R}^2$ be the function which stretches vectors in the $\vec{b}_1$ direction by 2 and leaves vectors in the $\vec{b}_2$ direction fixed. That is,

$$F(x_1\vec{b}_1 + x_2\vec{b}_2) = 2x_1\vec{b}_1 + x_2\vec{b}_2.$$

Find the defining matrices A_G and $A_{G,\mathcal{B}}$.

10.2. Matrix Similarity

Note that the defining matrix of a function depends on our choice of basis. Usually, if $\mathcal{B}$ and $\mathcal{C}$ are distinct bases for $\mathbb{R}^n$, then $A_{F,\mathcal{B}}$ and $A_{F,\mathcal{C}}$ will be different matrices. However, we don't want to lose track of the fact that these two matrices are related to each other. We define the following.

Definition 10.3 (Geometric definition of matrix similarity)**.** Two $n \times n$ matrices B and C are called *similar* if they represent the same function, but in possibly different bases. That is, there is a single linear function $F:\mathbb{R}^n \to \mathbb{R}^n$ so that

$$A_{F,\mathcal{B}} = B \text{ and } A_{F,\mathcal{C}} = C,$$

where $\mathcal{B}$ and $\mathcal{C}$ are bases for $\mathbb{R}^n$.

Activity 10.2. Let $B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Find a matrix C that's similar to B, but not equal to B.

Let's derive an algebraic method to detect matrix similarity using our results from the previous section. Let B and C be similar $n \times n$ matrices. Then, by definition, there exists a linear transformation $F:\mathbb{R}^n \to \mathbb{R}^n$ and bases $\mathcal{B}$, $\mathcal{C}$ for $\mathbb{R}^n$ so that

$$A_{F,\mathcal{B}} = B \text{ and } A_{F,\mathcal{C}} = C.$$

Recalling our definition of defining matrices from the previous section, we have

$$[F(\vec{x})]_{\mathcal{C}} = C[\vec{x}]_{\mathcal{C}}.$$

Now, let's use our change of basis matrix: we have $M_{\mathcal{C}\leftarrow\mathcal{B}}[F(\vec{x})]_{\mathcal{B}} = [F(\vec{x})]_{\mathcal{C}}$ and $M_{\mathcal{C}\leftarrow\mathcal{B}}[\vec{x}]_{\mathcal{B}} = [\vec{x}]_{\mathcal{C}}$. Replacing this into the equation above yields

$$M_{\mathcal{C}\leftarrow\mathcal{B}}[F(\vec{x})]_{\mathcal{B}} = CM_{\mathcal{C}\leftarrow\mathcal{B}}[\vec{x}]_{\mathcal{B}}.$$

For notational simplicity, let's write $P := M_{\mathcal{C}\leftarrow\mathcal{B}}$, so that we have

$$P[F(\vec{x})]_{\mathcal{B}} = CP[\vec{x}]_{\mathcal{B}}.$$

Recall that P is invertible, and so we can multiply the left-hand side of the equality above to get

$$[F(\vec{x})]_{\mathcal{B}} = P^{-1}CP[\vec{x}]_{\mathcal{B}}.$$

But then, by definition, $B = P^{-1}CP$. This yields the following equivalent definition of matrix similarity.

Definition 10.4 (Algebraic definition of matrix similarity)**.** Two $n \times n$ matrices B and C are called *similar* if there exists an invertible $n \times n$ matrix P so that

$$B = P^{-1}CP.$$

Activity 10.3. Show that if two matrices B and C are similar, then B is invertible if and only if C is invertible.

Activity 10.4. Show that the matrices

$$B = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \text{ and } C = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 1 & -1 & -1 \end{pmatrix}$$

are not similar.

Remark 10.5. With the tools we currently have, it is generally very difficult for us to determine whether two matrices are similar. In the following section, we develop a method to detect matrix similarity for certain special families of matrices.

10.3. Diagonalization

Let's return to our eigen-story from Chapter 9.

Activity 10.5. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$.

(1) Show that A has eigenvalues $\lambda = -1$ and $\lambda = 4$.
(2) Find vectors $\vec{v}$ and $\vec{w}$ so that $E_{-1} = \text{Span}(\vec{v})$ and $E_4 = \text{Span}(\vec{w})$.
(3) Observing that $\{\vec{v}, \vec{w}\}$ forms a basis for $\mathbb{R}^2$, show A is similar to the matrix $\begin{pmatrix} -1 & 0 \\ 0 & 4 \end{pmatrix}$.
(4) Give a geometric description for how the function T_A transforms $\mathbb{R}^2$.

Activity 10.6. Consider the matrix $B = \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix}$.

(1) Show that B has exactly one eigenvalue $\lambda = 2$, and that the geometric multiplicity of $\lambda = 2$ is equal to 1.
(2) Do you think it's possible for B to be similar to a matrix of the form
$$\begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}?$$
Explain why or why not.

We define the following.

Definition 10.6. A matrix is called DIAGONAL if the only nonzero entries in the matrix appear on the diagonal. We write
$$D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$$
to notate the diagonal matrix
$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ 0 & 0 & \cdots & d_n \end{pmatrix}.$$

Note that if $D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$ then
$$D \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} d_1 x_1 \\ d_2 x_2 \\ \vdots \\ d_n x_n \end{pmatrix},$$
and so the matrix transformation T_D can be understood as the transformation which stretches each coordinate of a vector by a of factor d_i, for $i = 1, \ldots, n$. We'll describe this as a DILATION TRANSFORMATION.

Definition 10.7. An $n \times n$ matrix A is called DIAGONALIZABLE if it is similar to a diagonal matrix.

Observe that a matrix A is diagonalizable if it's the defining matrix for a function $F : \mathbb{R}^n \to \mathbb{R}^n$ with respect to some basis $\mathcal{B}$ of $\mathbb{R}^n$. So, if we view $\mathbb{R}^n$ with respect to $\mathcal{B}$-coordinates, we see that F behaves as a dilation transformation.

The following result will help us characterize which matrices are diagonalizable.

Theorem 10.8 (The Diagonalization Theorem)**.** *An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. In this case, there are linearly independent eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ for A so that then $D = C^{-1}AC$ where*
$$D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \text{ and } C = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}.$$

Proof. Suppose first that A is diagonalizable. Then, by our algebraic definition of matrix similarity, there's an invertible matrix $C = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}$ and diagonal matrix $D = \operatorname{diag}(d_1, \ldots, d_n)$ so that

$$D = C^{-1}AC$$

Since C is invertible, we know that $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ is a linearly independent set, and hence forms a basis for $\mathbb{R}^n$. So, we just need to show that $\vec{v}_i$ is an eigenvector with eigenvalue d_i for all $i = 1, \ldots, n$. Observe that

$$C\vec{e}_i = \vec{v}_i \Rightarrow \vec{e}_i = C^{-1}\vec{v}_i \tag{10.1}$$

and so we have

$$\begin{aligned} A\vec{v}_i &= CDC^{-1}\vec{v}_i \\ &= CD\vec{e}_i, \text{ by Equation (10.1)} \\ &= Cd_i\vec{e}_i, \text{ since the } i\text{th column of } D \text{ is } d_i\vec{e}_i \\ &= d_iC\vec{e}_i \\ &= d_i\vec{c}_i, \text{ by Equation (10.1)} \end{aligned}$$

So, $\vec{v}_i$ is an eigenvector of A with eigenvalue d_i.

Conversely, suppose that A has n linearly independent eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$ corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$. Observe that $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ is a basis for $\mathbb{R}^n$ (note that a basis consisting of eigenvectors is often called an *eigenbasis*), and furthermore that the defining matrix for T_A with respect to our eigenbasis $\mathcal{B}$ is given by

$$D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

Now, using our derivation for the algebraic definition of matrix similarity gives

$$A = M_{\mathcal{E}\leftarrow\mathcal{B}}^{-1} D M_{\mathcal{E}\leftarrow\mathcal{B}}$$

and we have $M_{\mathcal{E}\leftarrow\mathcal{B}} = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}$, as needed. □

Activity 10.7. Determine which of the following matrices A_i are diagonalizable. For those that are, find an invertible matrix C_i and diagonal matrix D_i so that $D_i = C_i^{-1}A_iC_i$.

$$A_1 = \begin{pmatrix} 0 & 1 \\ 2 & 1 \end{pmatrix} A_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{pmatrix} A_3 = \begin{pmatrix} 4 & -3 & 0 \\ 2 & -1 & 0 \\ 1 & -1 & 1 \end{pmatrix}.$$

To use the Diagonalization Theorem, we need a method to determine whether A has enough linearly independent eigenvectors. We have the following Proposition, whose proof will be left as a Chapter Exercise.

Proposition 10.9. Let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues of a matrix A, and suppose that $\vec{v}_i \in E_{\lambda_i}$ for each $i \in \{1, \ldots, k\}$. Then $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is a linearly independent set.

This gives the following.

Corollary 10.10. If an $n \times n$ matrix A has n distinct eigenvalues, then A is diagonalizable.

Proof. If A has n distinct eigenvalues, then by Proposition 10.9 A must then have n distinct eigenvectors, and so A is diagonalizable by the Diagonalization Theorem. $\square$

Remark 10.11. Note that the converse of Proposition 10.9 does not hold. That is, it's not the case that if an $n \times n$ matrix A is diagonalizable then A must have n distinct eigenvalues. For example, if we let A be the matrix defined in Example 8.11, we see that A is diagonalizable but only has two eigenvalues. The following definition will help us characterize $n \times n$ diagonalizable matrices with less than n eigenvalues.

Lemma 10.12. *Let λ be an eigenvalue of a matrix A. Then, the geometric multiplicity is less than or equal to the algebraic multiplicity of λ.*

Proof. Suppose that λ is an eigenvalue of A with geometric multiplicity d, and suppose that $\{\vec{v}_1, \dots, \vec{v}_d\}$ is a basis for E_λ. By Chapter Exercise P3.9, we can extend this to a basis $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ for $\mathbb{R}^n$. Hence, the defining matrix for $F = T_A$ with respect to $\mathcal{B}$ is of the form

$$A_{F,\mathcal{B}} = \begin{pmatrix} \lambda I_d & X \\ 0 & Y \end{pmatrix}$$

for some matrices X and Y. By Chapter Exercise P10.**??**, we know that A and $A_{F,\mathcal{B}}$ have the same characteristic polynomial. So, it can be shown that

$$\chi_A(x) = (x - \lambda)^d \chi_Y(x).$$

Hence, the algebraic multiplicity of λ is greater than or equal to the geometric multiplicity d (with equality when λ is not an eigenvalue of Y). $\square$

We can now add the our Diagonalization Theorem. We have the following.

Theorem 10.13 (Diagonalization Theorem, final version)**.** *Let A be an $n \times n$ matrix whose eigenvalues are all real. The following are equivalent.*

(1) A is diagonalizable;

(2) The sum of the geometric multiplicities of A is equal to n;

(3) The geometric multiplicity of every eigenvalue λ is equal to the algebraic multiplicity of λ.

In this case, there are linearly independent eigenvectors $\vec{v}_1, \dots, \vec{v}_n$ with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ for A so that then $D = C^{-1}AC$ where

$$D = \operatorname{diag}(\lambda_1, \dots, \lambda_n) \text{ and } C = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}.$$

Note that, because of Proposition 10.9, (1) $\Leftrightarrow$ (2) can be seen as a rephrasing of our original Diagonalization Theorem (Theorem 10.8). All that's left to show in this final version is (2) $\Leftrightarrow$ (3).

Proof of Theorem 10.13. Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \dots, \lambda_k$ and let λ_i have geometric multiplicity d_i and algebraic multiplicity m_i. By the previous lemma, we know that $d_i \leq m_i$.

(2) $\Rightarrow$ (3): Suppose that $d_1 + \cdots + d_k = n$. Note that the characteristic polynomial of A is of degree n, and since we can write

$$\chi_A(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}$$

this gives $m_1 + \cdots + m_k = n$. So, if $d_i < m_i$ for any i we would have

$$d_1 + \cdots + d_k < m_1 + \cdots + m_k < n,$$

contradicting our assumption. Since by our Lemma $d_i \leq m_i$, we must then have the equality $d_i = m_i$ for every i.

(3) $\Rightarrow$ (2): Conversely, suppose that $m_i = d_i$ for all i. Then we have

$$n = m_1 + \cdots + m_k = d_1 + \cdots + d_k,$$

as desired. □

Remark 10.14. Note that the addition of part (3) to our Theorem doesn't quite simplify our computation when our matrix *is* diagonalizable: no matter what we do, we still need to compute the dimension of the λ-eigenspace for each value of λ. But it does give the potential to simplify our justification that a matrix *is not* diagonalizable: all we need to do is find ONE eigenvalue where the geometric and algebraic multiplicities do not agree.

Example 10.15. Consider the matrix

$$A_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{pmatrix}.$$

The characteristic polynomial of A_2 is given by

$$\chi_{A_2}(x) = -(x-2)(x-1)^2$$

and so $\lambda = 1$ has algebraic multiplicity 2. However,

$$A_2 - 1 \cdot I_3 = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 2 & -5 & 3 \end{pmatrix}$$

which is row equivalent to a matrix with 2 pivots. So, $\text{nullity}(A - I_3) = 1$ which tells us that the geometric multiplicity of $\lambda = 1$ is equal to 1. So, by our final version of the Diagonalization Theorem, we know that A_2 is not diagonalizable.

10.4. Eigendecompositions

Note that if A is diagonalizable with $D = C^{-1}AC$ we can write $A = CDC^{-1}$. This gives a useful decomposition of our matrix A. We have the following definition.

Definition 10.16. Suppose that A is an $n \times n$ diagonalizable matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding linearly independent eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$. Let $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $C = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}$. We call the equality

$$A = CDC^{-1}$$

the EIGENDECOMPOSITION of the matrix A.

Example 10.17. Let's find a transformation $F : \mathbb{R}^2 \to \mathbb{R}^2$ which stretches every vector in the $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ direction by 2 and in the $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$ direction by 3.

Observing that F is linear, we know that $F = T_A$ for a matrix A. Furthermore, we know that A has eigenvalues $2, 3$ with corresponding eigenvectors

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

By the Diagonalization Theorem, A has eigendecomposition

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

Note that the eigendecomposition can also help us compute large powers of a matrix. We have the following.

Proposition 10.18. Let A be a diagonalizable matrix with eigendecomposition $A = CDC^{-1}$. Then,

$$A^n = CD^nC^{-1}$$

for any integer n.

Proof. We have

$$\begin{aligned} A^n &= (CDC^{-1})^n \\ &= \underbrace{(CDC^{-1})(CDC^{-1}) \cdots (CDC^{-1})}_{n \text{ times}} \\ &= \underbrace{CD(C^{-1}C)DC^{-1} \cdots CDC^{-1}}_{n \text{ times}} \\ &= C \underbrace{DD \cdots D}_{n \text{ times}} C^{-1} \\ &= CD^nC^{-1}. \end{aligned}$$

□

Exercises

P10.1 In this problem, we'll prove Proposition 10.9.

(a) Let λ_1 and λ_2 be distinct eigenvalues of a matrix A, and suppose that $\vec{v}_1 \in E_{\lambda_1}$ and $\vec{v}_2 \in E_{\lambda_2}$. Show that if $x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{0}$ then

$$x_1\lambda_1\vec{v}_1 + x_2\lambda_2\vec{v}_2 = \vec{0} \text{ and } x_1\lambda_1\vec{v}_1 + x_2\lambda_1\vec{v}_2 = \vec{0}.$$

(b) Take the difference of the equalities above to show that $x_2 = 0$.

(c) Use a similar argument to show that $x_1 = 0$. Conclude that $\vec{v}_1, \vec{v}_2$ are linearly independent.

(d) Next, suppose that $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are eigenvectors of a matrix A with corresponding eigenvalues $\lambda_1, \lambda_2, \lambda_3$. If $\lambda_i \neq \lambda_j$ for any $i \neq j$ (that is, the λ_i are all distinct real numbers). Show that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a linearly independent set *(Hint: use a similar strategy to the previous parts!)*.

(e) Finally, suppose that $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ are eigenvectors of a matrix A with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. If $\lambda_i \neq \lambda_j$ for any $i \neq j$ (that is, the λ_i are all distinct real numbers), show that $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is a linearly independent set.

P10.2 Show that if A and B are similar matrices, then $\chi_A = \chi_B$. Conclude that similar matrices have the same eigenvalues.

P10.3 Show that if A is diagonalizable, then A^n is diagonalizable for any positive integer n.

P10.4 Show that if A is diagonalizable, then $A^\top$ is diagonalizable.

P10.5 Recall that a matrix A is called NILPOTENT if A^m is equal to the zero matrix for some positive integer m. True or False: If A is nilpotent and diagonalizable, then A must be equal to the zero matrix. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P10.6 True or False: If A is diagonalizable, then A must be invertible. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P10.7 A HORIZONTAL SHEAR is a function $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x + my \\ y \end{pmatrix}$$

and a VERTICAL SHEAR is a function $G : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$G\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x \\ y + mx \end{pmatrix}.$$

Show that the shear functions are not diagonalizable when $m \neq 0$.

P10.8 The *Fibonacci sequence* is the linear recurrence sequence $\{f_n\}$ defined by

$$f_0 = 0, f_1 = 1, \text{ and } f_{n+2} = f_{n+1} + f_n.$$

In this problem, we'll observe how eigendecompositions can help us study this sequence.

(a) Find f_{12} by hand (don't look it up, use the recurrence definition)

(b) Explain why the following identity holds for any integer n

$$\begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} f_{n-1} \\ f_n \end{pmatrix}$$

(c) Use the previous part to show that

$$\begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

for every integer n. (Note: you do not need to show this formally)

(d) Use the eigendecompositions to show that
$$f_n = \frac{\alpha^n - \beta^n}{\alpha - \beta}$$
where
$$\alpha = \frac{1+\sqrt{5}}{2} \text{ and } \beta = \frac{1-\sqrt{5}}{2}.$$
(Note that α, β are the eigenvalues of the matrix above).

(e) Use part (d) to compute f_{12} again and check that it matches what you found in part (a). Also compute f_{20}, f_{50} and f_{100}.

Chapter 11

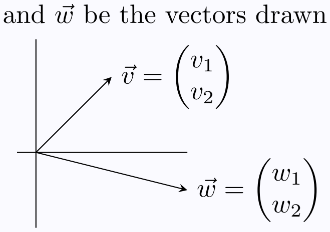
Orthogonality

11.1. The Dot Product

In this chapter, we'll see that knowing which vectors are perpendicular to each other can help us gain further understanding of how a linear transformation behaves. In this section, we develop algebraic machinery to detect this. Let's first consider the following example.

> **Activity 11.1.** Use geometric reasoning to show that the vectors $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ are perpendicular to each other.

In our activity above, in order to calculate the angle between our two vectors, we first needed to calculate the length of each vector and the distance between them. Let's look at these calculations more generally.

> **Activity 11.2.** Let $\vec{v}$ and $\vec{w}$ be the vectors drawn below.
>
> 
>
>
> Find a formula for the length of $\vec{v}$ and $\vec{w}$ and the distance between $\vec{v}$ and $\vec{w}$.

To generalize notions of distance and angle, we'll define the following operation.

Definition 11.1. Let

$$\vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \text{ and } \vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

be vectors in $\mathbb{R}^n$. The DOT PRODUCT of $\vec{u}$ and $\vec{v}$ is the scalar

$$\vec{u} \cdot \vec{v} := u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Note that is we think of the vector $\vec{u}$ as an $n \times 1$ matrix, we can define the dot product instead as the matrix-vector product $\vec{u} \cdot \vec{v} = \vec{u}^\top \vec{v}$.

Observe that the dot product satisfies the following properties.

Proposition 11.2. Let $\vec{u}, \vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^n$ and let c be a scalar. Then,

(1) *Commutativity:* $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
(2) *Distributivity with Addition:* $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$
(3) *Distributivity with Scalar Multiplication:* $(c\vec{u}) \cdot \vec{v} = c(\vec{u} \cdot \vec{v})$

Observation 11.3. Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^2$. Note that the length of $\vec{u}$ is equal to $\sqrt{\vec{u} \cdot \vec{u}}$ and the distance between $\vec{u}$ and $\vec{v}$ is given by $\sqrt{(\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v})}$. Let's use this observation to extend our notions of distance to n-dimensions.

Definition 11.4. The NORM of a vector $\vec{u}$ in $\mathbb{R}^n$ is defined by

$$\|\vec{u}\| := \sqrt{\vec{u} \cdot \vec{u}}.$$

Definition 11.5. Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^n$. Then, the DISTANCE between $\vec{u}$ and $\vec{v}$, denoted $d(\vec{u}, \vec{v})$ is equal to the length of the vector $\vec{u} - \vec{v}$. That is,

$$d(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\|.$$

Now, let's generalize our work from Activity 11.1 to any two vectors in $\mathbb{R}^2$.

Definition 11.6. Let $\vec{u}$ and $\vec{v}$ be nonparallel vectors in $\mathbb{R}^n$. Then, the ANGLE between vectors $\vec{u}$ and $\vec{v}$ is the smaller of the two angles between them. If $\vec{u}$ and $\vec{v}$ are parallel, then the angle between them is 0.

Activity 11.3. Let $\vec{u}$ and $\vec{v}$ be any two vectors in $\mathbb{R}^2$. In this activity, we'll derive a formula for the angle between $\vec{u}$ and $\vec{v}$.

(1) Show that for any vector $\vec{x}$ in $\mathbb{R}^n$ we have $\|\vec{x}\|^2 = \vec{x} \cdot \vec{x}$.
(2) Use law of cosines to show that the angle between $\vec{u}$ and $\vec{v}$ is given by

$$\theta = \arccos \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \|\vec{v}\|}.$$

(Hint: use the observation in part (1), along with Proposition 11.2)

(3) Conclude that $\vec{u}$ and $\vec{v}$ are perpendicular if and only if $\vec{u} \cdot \vec{v} = 0$.

We can now use our work in $\mathbb{R}^2$ to define the angle between higher-dimensional vectors. We have the following.

Definition 11.7. Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^n$. Then, the ANGLE between $\vec{u}$ and $\vec{v}$ is defined by

$$\theta = \arccos \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \|\vec{v}\|}.$$

We say that $\vec{u}$ and $\vec{v}$ are ORTHOGONAL if $\vec{u} \cdot \vec{v} = 0$.

Remark 11.8. Note that this definition matches our notions of geometry in $\mathbb{R}^3$ as well. In fact, our derivation for the angle between two three-dimensional vectors would follow identically to our work in Activirty 11.3.

Activity 11.4. Determine which of the following pairs of vectors $\vec{u}$ and $\vec{v}$ are orthogonal.

(1) $\vec{u} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$

(2) $\vec{u} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$

(3) $\vec{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$

11.2. Orthonormal Bases and Orthogonal Matrices

In the previous chapters, we saw that the fundamental object needed to understand a vector space is a basis. We learned that real vector spaces of dimension n are all isomorphic to $\mathbb{R}^n$, and we saw how different bases define coordinate systems on our vector spaces which can help us better understand certain linear transformations. In this section, we look at how bases interact with the dot product.

Definition 11.9. A basis $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ is ORTHOGONAL if $\vec{v}_i \cdot \vec{v}_j = 0$ for every $i \neq j$. If it's also the case that $\|\vec{v}_i\| = 1$ for every i we call $\mathcal{B}$ an ORTHONORMAL BASIS for $\mathbb{R}^n$.

Activity 11.5. Consider the bases $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ for $\mathbb{R}^2$ given by

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}, \mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \text{ and } \mathcal{D} = \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right\},$$

(1) Determine which of the bases above are orthogonal and which are orthonormal.

(2) Calculate $\vec{u} \cdot \vec{u}$ given that $[\vec{u}]_{\mathcal{B}} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$.

(3) Calculate $\vec{v} \cdot \vec{v}$ given that $[\vec{v}]_{\mathcal{C}} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$.

(4) Calculate $\vec{w}\cdot\vec{w}$ given that $[\vec{w}]_{\mathcal{D}} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$.

(5) What did you notice in your calculations?

The following Proposition tells us that orthonormal bases preserve dot products, and as a consequence preserve distances and angles.

Proposition 11.10. Let $\mathcal{B}$ be an orthonormal basis for $\mathbb{R}^n$ and take any vectors $\vec{x}, \vec{y}$ in $\mathbb{R}^n$. Then

$$[\vec{x}]_{\mathcal{B}} \cdot [\vec{y}]_{\mathcal{B}} = \vec{x} \cdot \vec{y}.$$

In particular, we have $\|\vec{x}\| = \|[\vec{x}]_{\mathcal{B}}\|$.

Proof. Suppose that $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$ and write

$$[\vec{x}]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ and } [\vec{y}]_{\mathcal{B}} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

That is,

$$\vec{x} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n$$
$$\vec{y} = y_1\vec{v}_1 + \cdots + y_n\vec{v}_n.$$

Then we have

$$\vec{x}\cdot\vec{y} = (x_1\vec{v}_1 + \cdots + x_n\vec{v}_n) \cdot (y_1\vec{v}_1 + \cdots + y_n\vec{v}_n)$$

Using the distributive properties of the dot product (Proposition 11.2), we'll end up with a sum of terms of the form

$$x_i y_j \vec{v}_i \vec{v}_j.$$

But, since we know that $\vec{v}_i \cdot \vec{v}_j = 0$ whenever $i \neq j$ then we have

$$\vec{x}\cdot\vec{y} = x_1 y_1 \vec{v}_1 \cdot \vec{v}_1 + \cdots x_n y_n \vec{v}_n \cdot \vec{v}_n.$$

But we also know that $\vec{v}_i \cdot \vec{v}_i = \|\vec{v}\|^2 = 1$, since our basis is orthonormal. So, we have

$$\vec{x}\cdot\vec{y} = x_1 y_1 + \cdots + x_n y_n$$

as desired. $\square$

Activity 11.6. Suppose that $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}$ is an orthonormal basis for $\mathbb{R}^2$, and consider the matrix $Q = \begin{pmatrix} \vec{b}_1 & \vec{b}_2 \end{pmatrix}$. Show that Q is invertible with $Q^{-1} = Q^\top$.

More generally, we have the following.

Proposition 11.11. Let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a basis for $\mathbb{R}^n$ and let Q be the matrix with column vectors $\vec{v}_1, \ldots, \vec{v}_n$. Then $\mathcal{B}$ is orthonormal if and only if $Q^{-1} = Q^\top$.

Proof. Let Q be the matrix with column vectors $\vec{v}_i$. Then $Q^\top$ is the matrix with *rows* $\vec{v}_i$, and so we can observe that

$$Q^\top Q = \begin{pmatrix} \vec{v}_1 \cdot \vec{v}_1 & \vec{v}_1 \cdot \vec{v}_2 & \cdots & \vec{v}_1 \cdot \vec{v}_n \\ \vec{v}_2 \cdot \vec{v}_1 & \vec{v}_2 \cdot \vec{v}_2 & \cdots & \vec{v}_2 \cdot \vec{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{v}_n \cdot \vec{v}_1 & \vec{v}_n \cdot \vec{v}_2 & \cdots & \vec{v}_n \cdot \vec{v}_n \end{pmatrix}.$$

So, $Q^\top Q = I_n$ if and only if

$$\vec{v}_i \cdot \vec{v}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

which occurs precisely when $\{\vec{v}_1, \ldots, \vec{v}_n\}$ forms an orthonormal basis for $\mathbb{R}^n$. □

This gives rise to the following (somewhat annoying) definition.

Definition 11.12. We call a matrix Q ORTHOGONAL if $Q^{-1} = Q^\top$. Equivalently, Q is called orthogonal if its column vectors form an orthonormal basis.

Remark 11.13. This definition is annoying, because orthogonal matrices aren't just those matrices with *orthogonal* column vectors, but rather with *orthonormal* column vectors. I don't know why we don't just call them orthonormal matrices. My guess is because matrices with column vectors that are orthogonal, but not orthonormal, don't have many nice properties so they don't get their own name.

Just for fun and in case you're interested, Hadamard matrices are matrices with orthogonal (but not orthonormal) column vectors which only have entries equal to ± 1. It can be shown that for an $n \times n$ Hadamard matrix H we have $HH^\top = nI_n$ so that $H^{-1} = (1/n)H^\top$.

As a consequence of Proposition 11.10, we can show that orthogonal matrices preserve distances and angles. We have the following.

Theorem 11.14. *Let Q be an $n \times n$ orthogonal matrix. Then, for any $\vec{v}, \vec{w}$ in $\mathbb{R}^n$*

$$Q\vec{v} \cdot Q\vec{w} = \vec{v} \cdot \vec{w}.$$

In particular, $\|Q\vec{v}\| = \|\vec{v}\|$ and if θ is the angle between $\vec{v}$ and $\vec{w}$ then θ is also the angle between $Q\vec{v}$ and $Q\vec{w}$.

Proof. Let $Q = \begin{pmatrix} \vec{v}_1 & \cdots \vec{v}_n \end{pmatrix}$, and observe that Q is the change of basis matrix $Q = M_{\mathcal{E} \leftarrow \mathcal{B}}$ where $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$. So,

$$Q\vec{v} \cdot Q\vec{w} = [\vec{v}]_{\mathcal{B}} \cdot [\vec{w}]_{\mathcal{B}} = \vec{v} \cdot \vec{w},$$

where the final equality follows by Proposition 11.10. This proves our claim.

Note that the could also prove this algebraically – we include this version as well for completeness. We have

$$\begin{aligned} Q\vec{v} \cdot Q\vec{w} &= (Q\vec{v})^\top Q\vec{w} \\ &= \vec{v}^\top Q^\top Q\vec{w} \\ &= \vec{v}^\top \vec{w}, \text{ since } Q^\top = Q^{-1} \\ &= \vec{v} \cdot \vec{w}, \end{aligned}$$

as needed. $\square$

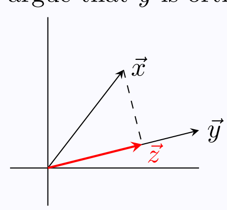
Remark 11.15. Theorem 11.14 tells us that when Q is orthogonal, T_Q is a transformation that preserves distances and angles. So, we can reason geometrically that Q must be either a rotation or reflection transformation. In Chapter Exercise P11.11, you'll prove this formally in the 2×2 case.

11.3. The Gram-Schmidt Process

In this section, we'll show that every vector space has an orthonormal basis, and in fact that we can produce an orthonormal basis algorithmically given any generating set for our space. To do this, we'll need to develop the notion of "orthogonal projections".

Activity 11.7. Let $\vec{x}$ and $\vec{y}$ be vectors in $\mathbb{R}^2$, and let $\vec{z}$ be the closets point in Span($\vec{y}$) to $\vec{x}$.

(1) Use the picture below to argue that $\vec{y}$ is orthogonal to $\vec{x} - \vec{z}$.



(2) Since $\vec{z}$ is in Span($\vec{y}$), we can write $\vec{z} = c\vec{y}$ for some real number c. Use the previous part to show that

$$c = \frac{\vec{x} \cdot \vec{y}}{\vec{y} \cdot \vec{y}}.$$

(3) Conclude that the closest point on Span($\vec{y}$) to $\vec{x}$ is given by

$$\vec{z} = \frac{\vec{x} \cdot \vec{y}}{\vec{y} \cdot \vec{y}}\vec{y}.$$

Using our work in $\mathbb{R}^2$, we can define the following.

Definition 11.16. For vectors $\vec{x}, \vec{y}$ in $\mathbb{R}^n$, the ORTHOGONAL PROJECTION of $\vec{x}$ onto $\vec{y}$ is given by

$$\operatorname{proj}_{\vec{y}} \vec{x} := \frac{\vec{x} \cdot \vec{y}}{\vec{y} \cdot \vec{y}}\vec{y}.$$

As we saw in Activity 11.7, when our vectors are in $\mathbb{R}^2$ or $\mathbb{R}^3$, the orthogonal projection $\text{proj}_{\vec{y}}\,\vec{x}$ is the closest point on $\text{Span}(\vec{y})$ to $\vec{x}$. We are now prepared to prove the following.

Theorem 11.17 (The Gram-Schmidt Process)**.** *Every vector space has an orthogonal basis. Furthermore, if V is a vector subspace of $\mathbb{R}^n$ with basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$, and we let*

$$\begin{aligned}
\vec{u}_1 &= \vec{v}_1 \\
\vec{u}_2 &= \vec{v}_2 - \mathit{proj}_{\vec{u}_1}\,\vec{v}_2 \\
\vec{u}_3 &= \vec{v}_3 - \mathit{proj}_{\vec{u}_1}\,\vec{v}_3 - \mathit{proj}_{\vec{u}_2}\,\vec{v}_3 \\
&\vdots \\
\vec{u}_m &= \vec{v}_m - \mathit{proj}_{\vec{u}_1}\,\vec{v}_m - \mathit{proj}_{\vec{u}_2}\,\vec{v}_m - \cdots - \mathit{proj}_{\vec{u}_{m-1}}\,\vec{v}_m,
\end{aligned}$$

then, $\{\vec{u}_1, \ldots, \vec{u}_m\}$ is an orthogonal basis for V.

The proof for Gram-Schmidt needs induction. Instead of working out these technical details, let's take a look at an example to understand why the result holds.

Example 11.18. Let $V = \text{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ be the subspace of $\mathbb{R}^3$ with

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \text{ and } \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}.$$

Note that $V = \mathbb{R}^3$ in this case, and we know that the standard basis forms an orthonormal basis, but let's look at how Gram-Schmidt edits the given basis.

First, set $\vec{u}_1 := \vec{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$. Then we let

$$\vec{u}_2 := \vec{v}_2 - \text{proj}_{\vec{u}_1}\,\vec{v}_2 = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}.$$

Observe that $\vec{u}_1$ is orthogonal to $\vec{u}_2$. We can see why this step works geometrically (as in part (1) of Activity 11.7), or algebraically as demonstrated below

$$\vec{u}_1 \cdot \vec{u}_2 = \vec{u}_1 \left(\vec{v}_2 - \frac{\vec{v}_2 \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1 \right) = \vec{u}_1 \cdot \vec{v}_2 - \frac{\vec{v}_2 \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1 \cdot \vec{u}_1 = 0.$$

Furthermore, since $\text{proj}_{\vec{u}_1}\,\vec{v}_2 \in \text{Span}(\vec{u}_1) = \text{Span}(\vec{v}_1)$ then we can write $\vec{u}_2 = \vec{v}_2 - c\vec{v}_1$ for some $c \in \mathbb{R}$ and so $\vec{u}_2 \in V$. Next, we let

$$\vec{u}_3 := \vec{v}_3 - \text{proj}_{\vec{u}_1}\,\vec{v}_3 - \text{proj}_{\vec{u}_2}\,\vec{v}_3 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}.$$

Observe that $\vec{u}_3$ is orthogonal to $\vec{u}_1$ and $\vec{u}_2$. We can see why this step works by noting the following:

- $\vec{u}_1$ is orthogonal to $\vec{v}_3 - \text{proj}_{\vec{u}_1}\,\vec{v}_3$, using an argument similar to above;

- Since $\vec{u}_1$ is orthogonal to $\vec{u}_2$ from the previous step, and $\operatorname{proj}_{\vec{u}_2} \vec{v}_3$ is in $\operatorname{Span}(\vec{u}_2)$, then $\vec{u}_1$ is orthogonal to $\operatorname{proj}_{\vec{u}_2} \vec{v}_3$.

Combining these observations gives

$$\begin{aligned}\vec{u}_1 \cdot \vec{u}_3 &= \vec{u}_1 \cdot (\vec{v}_3 - \operatorname{proj}_{\vec{u}_1} \vec{v}_3 - \operatorname{proj}_{\vec{u}_2} \vec{v}_3) \\ &= \vec{u}_1 \cdot (\vec{v}_3 - \operatorname{proj}_{\vec{u}_1} \vec{v}_3) - \vec{u}_1 \cdot \operatorname{proj}_{\vec{u}_2} \vec{v}_3) \\ &= 0 - 0 = 0.\end{aligned}$$

We can argue similarly to show that $\vec{u}_2 \cdot \vec{u}_3 = 0$ and so $\mathcal{B} = \{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ is an orthogonal set. Now, we need to check that it's still a basis for V. By Chapter Exercise P11.12, the set is linearly independent, and from above we know that $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ is a subset of V. Since $\dim(V) = 3$, we must have that $\mathcal{B}$ forms a basis, as needed (since if the set didn't generate, then by Chapter Exercise P3.8 we could extend this to a basis with more than 3 elements).

Finally, now that we have an orthogonal set, we can make it *orthonormal* by dividing through by the norm. In our example, we obtain the orthonormal basis

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

The key point we should take away from this section is that **every vector space has an orthonormal basis**, and moreover that such a basis can be found algorithmically as demonstrated above.

11.4. The Spectral Theorem

We define the following special family of matrices.

Definition 11.19. An $n \times n$ matrix A is ORTHOGONALLY DIAGONALIZABLE if there exists an orthogonal matrix Q and a diagonal matrix D so that $Q^\top A Q = D$.

Remark 11.20. Note that an $n \times n$ matrix A being orthogonally diagonalizable is equivalent to the existence of an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A. This is a particularly nice situation, since orthonormal bases preserve the dot product, and bases consisting of eigenvectors help us understand the linear transformation T_A. The following result completely characterizes when we're in this situation.

Theorem 11.21 (The Spectral Theorem)**.** *An $n \times n$ matrix A is orthogonally diagonalizable if and only if it is symmetric (that is, $A = A^\top$).*

We first need the following lemma.

Lemma 11.22. *Let A be symmetric. Then,*

(1) A has at least one real eigenvalue, and

(2) if λ, μ are distinct eigenvalues of A, then for any $\vec{x} \in E_\lambda$ and $\vec{y} \in E_\mu$ we have that $\vec{x}$ and $\vec{y}$ are orthogonal.

Proof. By the Fundamental Theorem of Algebra (Theorem 8.8), we know that A has at least one *complex* eigenvalue λ and corresponding eigenvector $\vec{v}$. We have

$$\begin{aligned} A\vec{v} \cdot A\vec{v} &= (A\vec{v})^\top (A\vec{v}) \\ &= \vec{v}^\top A^\top A\vec{v} \\ &= \vec{v}^\top A^2 \vec{v} \\ &= \vec{v}^\top \lambda^2 \vec{v} \\ &= \lambda^2 \vec{v} \cdot \vec{v} \\ &= \lambda^2 \|\vec{v}\|. \end{aligned}$$

Hence, $\lambda^2 = \|A\vec{v}\|^2 / \|\vec{v}\| > 0$ and so λ is real.

Next, suppose that A has eigenvalues $\lambda \neq \mu$ and let $\vec{x} \in E_\lambda$ and $\vec{y} \in E_\mu$. Then we have

$$\begin{aligned} \lambda \vec{x} \cdot \vec{y} &= A\vec{x} \cdot \vec{y} \\ &= (A\vec{x})^\top \vec{y} \\ &= \vec{x}^\top A^\top \vec{y} \\ &= \vec{x}^\top A\vec{y}, \text{ since } A = A^\top \\ &= \vec{x} \cdot \mu\vec{y} \\ &= \mu\vec{x} \cdot \vec{y}, \end{aligned}$$

and so $(\lambda - \mu)\vec{x} \cdot \vec{y} = 0 \Rightarrow \vec{x} \cdot \vec{y} = 0$ since we've assumed that $\lambda \neq \mu$. □

We are now prepared to prove the Spectral Theorem.

Proof of Theorem 11.21. Suppose first that A is orthogonally diagonalizable. Then, there exists an orthogonal matrix Q and diagonal matrix D so that

$$Q^\top A Q = D \Rightarrow A = QDQ^\top.$$

This gives

$$A^\top = (Q^\top)^\top D^\top Q^\top = QDQ^\top = A$$

noting that diagonal matrices are symmetric.

Conversely, suppose that A is symmetric. By Lemma 11.22 we know that A has a real eigenvalue, and that eigenvectors of A with distinct eigenvalues are orthogonal. Furthermore, by Gram-Schmidt we know that we can find an orthonormal basis for each λ-eigenspace. So, what's left to show is that our matrix is diagonalizable. To do this, we'll need proof by induction. We sketch the rest of this proof below for completeness, but note that this method is beyond the scope of this course.

Note that any 1×1 matrix is diagonalizable. So, suppose that A is $n \times n$ and symmetric. Let λ be a real eigenvalue of A with eigenvector $\vec{v}_1$. By Gram-Schmidt, we can extend this to an orthonormal basis $\{\vec{v}_1, \ldots, \vec{v}_n\}$ for $\mathbb{R}^n$. So, if we let

$$Q = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix},$$

then Q is orthogonal. Observe that $Q^\top AQ$ is of the form

$$Q^\top AQ = \begin{pmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & A_1 \end{pmatrix},$$

where A_1 is $(n-1)\times(n-1)$. Observing that A_1 is symmetric, by induction it's diagonalizable, and so the result follows. $\square$

The Spectral Theorem tells us that any symmetric matrix A has a SPECTRAL DECOMPOSITION. That is, we can write

$$A = QDQ^\top$$

for an orthogonal matrix Q and diagonal matrix D. So, transformations defined by a symmetric matrix must look like some rotation/reflection, followed by a dilation (multiplying by a diagonal matrix), and then followed by the opposite rotation/reflection.

Activity 11.8. Consider the matrix $A = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}$.

(1) Check that this matrix is symmetric.
(2) Verify that $\chi_A(x) = (7-x)(2-x)$ and that

$$E_7 = \mathrm{Span}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right) \text{ and } E_2 = \mathrm{Span}\left(\begin{pmatrix} -2 \\ 1 \end{pmatrix}\right).$$

(3) Find an ortho*normal* basis of eigenvectors for A.
(4) Find a spectral deomposition for A. That is, find an orthogonal matrix Q and diagonal matrix A so that $A = QDQ^\top$.
(5) Since Q is a rotation matrix, we can write

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Find the value of θ satisfying the equality above.
(6) Using your work in the previous parts, give a complete geometric description of how T_A transforms $\mathbb{R}^2$.

While the Spectral Theorem might seem like a special edge case, we can actually use this result to obtain a similar geometric understanding of *any* matrix transformation. We explore this in our final section of the semester.

11.5. The Singular Value Decomposition

In this section, we'll derive an important decomposition for *any* $m \times n$ matrix. Geometrically, this decomposition will show as that any linear transformation can be decomposed into a composition of three transformations: a rotation/reflection, followed by a dilation, followed by another rotation/reflection (not necessarily inverse to the original rotation/reflection).

This decomposition will rest on the following result.

Proposition 11.23. Let A be an $m \times n$ matrix. Then, there exists an orthonormal basis for $\mathbb{R}^n$ of eigenvectors of $A^\top A$ so that $\{A\vec{v}_1, \ldots, A\vec{v}_n\}$ is an orthogonal subset of $\mathbb{R}^m$. Furthermore, if we reindex our basis so that $A\vec{v}_1, \ldots, A\vec{v}_r$ are nonzero, and $A\vec{v}_{r+1} = \cdots = A\vec{v}_n = \vec{0}$, then $\{A\vec{v}_1, \ldots, A\vec{v}_r\}$ forms an orthogonal basis for $\mathrm{Col}(A)$.

Proof. Observe first that $A^\top A$ is symmetric. Indeed, we have

$$(A^\top A)^\top = A^\top (A^\top)^\top = A^\top A.$$

So, by the Spectral Theorem, there there exists an orthonormal basis for $\mathbb{R}^n$ of eigenvectors of $A^\top A$. Suppose that the eigenvectors $\vec{v}_i$ of $A^\top A$ have corresponding eigenvalues λ_i. Then, for any $i \neq j$ we have

$$\begin{aligned}
(A\vec{v}_i) \cdot (A\vec{v}_j) &= (A\vec{v}_i)^\top (A\vec{v}_j) \\
&= \vec{v}_i^\top A^\top A\vec{v}_j \\
&= \vec{v}_i^\top (\lambda_j \vec{v}_j) \\
&= \vec{v}_i \cdot (\lambda_j \vec{v}_j) \\
&= \lambda_j (\vec{v}_i \cdot \vec{v}_j) \\
&= 0,
\end{aligned}$$

where the final equality follows because $\vec{v}_i$ and $\vec{v}_j$ are orthogonal when $i \neq j$. Hence, $\{A\vec{v}_1, \ldots, A\vec{v}_m\}$ is an orthogonal subset of $\mathbb{R}^m$.

Next, if we reindex as in the theorem statement, we see that $A\vec{y} \in \mathrm{Col}(A)$ if and only if $A\vec{y} = A(x_1\vec{v}_1 + \cdots + x_n\vec{v}_n) = x_1 A\vec{v}_1 + \cdots + x_r A\vec{v}_r + \vec{0}$, and so

$$\mathrm{Col}(A) = \mathrm{Span}(A\vec{v}_1, \ldots, A\vec{v}_r).$$

Since $\{A\vec{v}_1, \ldots, A\vec{v}_r\}$ is orthogonal, then by Chapter Exercise P11.12 this set is linearly independent, and hence is a basis as needed. □

Let's look at an example to see what this Proposition gives us.

Example 11.24. Consider the 3×2 matrix

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

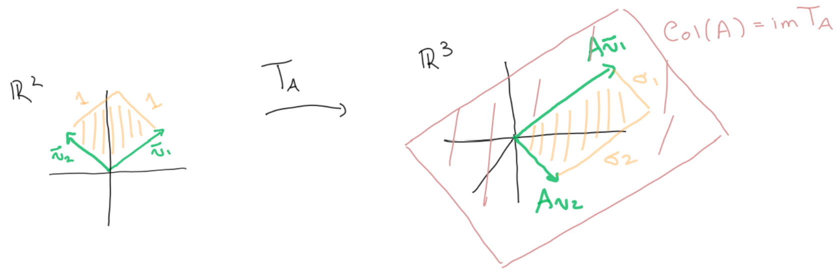
Then

$$A^\top A = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}.$$

By Activity 11.8 we know that $A^\top A$ has an orthonormal basis of eigenvectors $\{\vec{v}_1, \vec{v}_2\}$ with corresponding eigenvalues $\lambda_1 = 7$ and $\lambda_2 = 2$ where

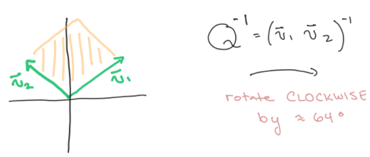
$$\vec{v}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \text{ and } \vec{v}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Recall that we also showed the orthogonal matrix $Q = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 \end{pmatrix}$ rotates vectors in $\mathbb{R}^2$ roughly 64° counterclockwise. Finally, note that by Proposition 11.27 we know that $\{A\vec{v}_1, A\vec{v}_2\}$ forms an orthogonal basis for $\mathrm{Col}(A)$. Putting this all together, we can see that $T_A : \mathbb{R}^2 \to \mathbb{R}^3$ is the transformation which sends the orthonormal basis $\{\vec{v}_1, \vec{v}_2\}$ to the orthogonal basis $\{A\vec{v}_1, A\vec{v}_2\}$ for $\mathrm{Col}(A)$ as pictured below

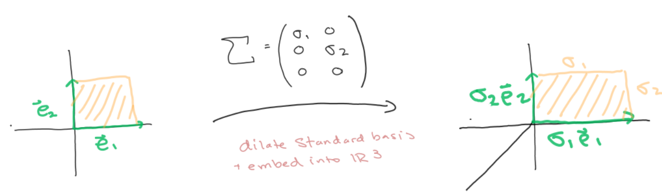


Next, let's observe how we could break up this transformation into a few steps. For convenience, let $\sigma_1 = \|A\vec{v}_1\|$ and $\sigma_2 = \|A\vec{v}_2\|$.

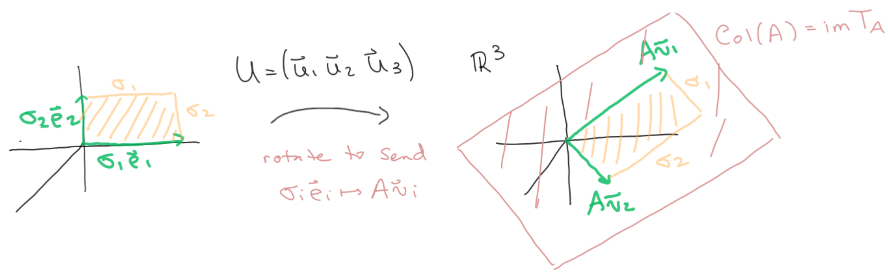
(1) First, rotate $\mathbb{R}^2$ clockwise by roughly $64°$ to send $\vec{v}_i \mapsto \vec{e}_i$ via the transformation defined by Q^{-1}



(2) Next, dilate the standard basis so that $\vec{e}_i \mapsto \sigma_i\vec{e}_i$, and then embed $\mathbb{R}^2 \hookrightarrow \mathbb{R}^3$ by placing $\mathbb{R}^2$ in the xy-plane. Observe that this step is accomplished via the transformation defined by $\Sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{pmatrix}$



(3) Finally, we can rotate $\mathbb{R}^3$ to send $\sigma_i\vec{e}_i \mapsto A\vec{v}_i$.



It takes a little more work to find this defining matrix. Let $\vec{u}_i = \frac{A\vec{v}_i}{\sigma_i}$. Recall that we defined $\sigma_i = \|A\vec{v}_i\|$ and so the vectors $\vec{u}_i$ have length 1. Furthermore, by Proposition 11.27 we know that $A\vec{v}_1$ and $A\vec{v}_2$ are orthogonal, and so $\vec{u}_1, \vec{u}_2$ is orthogonal as well. By Gram Schmidt, we can extend to an orthogonal basis $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ for $\mathbb{R}^3$. Hence, the orthogonal matrix $U = \begin{pmatrix} \vec{u}_1 & \vec{u}_2 & \vec{u}_3 \end{pmatrix}$ rotates $\mathbb{R}^3$ to send $\sigma_i \vec{e}_i \mapsto A\vec{v}_i$ as needed.

Summarizing what we've done above, and working out all of the necessary calculations, we see that our linear transformation can be understood by: (1) rotating $\mathbb{R}^2$ roughly 64° clockwise, (2) stretching $\vec{e}_1$ by $\sigma_1 = \sqrt{7}$ and $\vec{e}_2$ by $\sigma_2 = \sqrt{2}$ and then embedding $\mathbb{R}^2$ into $\mathbb{R}^3$, and finally (3) rotating $\mathbb{R}^3$ via the matrix U. Note that we could even spend a little more time to analyze the rotation matrix U (we could find its axis of rotation to be in the direction of its eigenvector, and there are shortcuts to find the angle we rotate by). But since we have limited time together, we'll just leave it at the fact that U is some rotation matrix.

The singular value decomposition guarantees that *every* linear transformation can be understood in this way. We have the following.

Theorem 11.25 (The Singular Value Decomposition)**.** *Let A be an $m \times n$ matrix. Then, there exists an orthogonal $m \times m$ matrix U, an orthogonal $n \times n$ matrix Q and a "block diagonal" matrix Σ so that*

$$A = U\Sigma V^\top.$$

(Note that in the notation from our previous example, we would have $V = Q$. We change notation in the statement above to be consistent with the literature).

Proof. Suppose that A is an $m \times n$ matrix, and let $\{\vec{v}_1, \ldots, \vec{v}_n\}$ be the eigenvalues of $A^\top A$ which form an orthonormal basis for $\mathbb{R}^n$. By Proposition 11.11 we know that

$$V = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}$$

is an orthogonal matrix. Furthermore, we have

$$AV = \begin{pmatrix} A\vec{v}_1 & \cdots & A\vec{v}_n \end{pmatrix}.$$

By Proposition 11.27 if we set $A\vec{v}_i = \sigma_i \vec{u}_i$, then (up to reindexing) we know that

$$\{\vec{u}_1, \ldots, \vec{u}_r\}$$

forms an orthonormal basis for $\mathrm{Col}(A)$. By Gram-Schmidt, we can extend this to an orthonormal basis of $\mathbb{R}^m$, say $\{\vec{u}_1, \ldots, \vec{u}_m\}$. Since we have $\vec{sigma}_i = 0$ for $i = r+1, \ldots, n$ we get that $A\vec{v}_i = \sigma_i \vec{u}_i$ for all i. We have a few cases.

If $n = m$, then we have

$$AV = \begin{pmatrix} \sigma_1 \vec{u}_1 & \cdots & \sigma_n \vec{u}_n \end{pmatrix} = U\Sigma,$$

where

$$U = \begin{pmatrix} \vec{u}_1 & \cdots & \vec{u}_n \end{pmatrix} \text{ and } \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n).$$

Next, if $n > m$ then we have

$$AV = \begin{pmatrix} A\vec{v}_1 & \cdots & A\vec{v}_n \end{pmatrix} = U\Sigma$$

where

$$U = \begin{pmatrix} \vec{u}_1 & \cdots & \vec{u}_m \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m & 0 & \cdots & 0 \end{pmatrix}$$

where there are an additional $n - m$ columns of zeros at the end of the matrix Σ.

Finally, if $n < m$ then we have

$$AV = \begin{pmatrix} A\vec{v}_1 & \cdots & A\vec{v}_n \end{pmatrix} = U\Sigma$$

where

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

where there are an additional $n - m$ rows of zeros at the bottom of the matrix Σ.

□

We define the following.

Definition 11.26. Let A be an $m \times n$ matrix and $\vec{v}_1, \ldots, \vec{v}_n$ be an orthonormal basis for $\mathbb{R}^n$ of eigenvectors for $A^\top A$, as above. The SINGULAR VALUES of A are given by $\sigma_i := \|A\vec{v}_i\|$.

Given our work with the singular value decomposition, we can extract geometric information by calculating the singular values of a matrix. The following Proposition gives a more practical method to calculate these values.

Proposition 11.27. Let A be an $m \times n$ matrix and $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $A^\top A$. Then, $\lambda_i > 0$ and the singular values of A are given by $\sigma_i = \sqrt{\lambda_i}$.

Proof. We have

$$\begin{aligned}
\sigma_i^2 &= \|A\vec{v}_i\|^2 \\
&= (A\vec{v}_i) \cdot (A\vec{v}_i) \\
&= (A\vec{v}_i)^\top (A\vec{v}_i) \\
&= \vec{v}_i^\top A^\top A\vec{v}_i \\
&= \vec{v}_i^\top \lambda_i \vec{v}_i \\
&= \lambda_i \vec{v}_i \cdot \vec{v}_i \\
&= \lambda_i \|\vec{v}_i\|^2 \\
&= \lambda_i,
\end{aligned}$$

where the final equality follows because the $\vec{v}_i$ form an ortho**normal** set. Hence, $\sigma_i^2 = \lambda_i$ and so $\lambda_i > 0$ and $\sigma_i = \sqrt{\lambda_i}$. ☐

Activity 11.9. Find the singular values of the following matrices. Given your calculations, what can you say about the corresponding transformations?

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

(2) $B = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{pmatrix}$

(3) $C = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

Exercises

P11.1 True or False: If $\vec{x}$ is a vector in $\mathbb{R}^n$ and $\vec{x} \cdot \vec{x} = 0$ then $\vec{x} = \vec{0}$. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P11.2 True or False: If $\vec{x}$ and $\vec{y}$ are vectors in $\mathbb{R}^n$ then $\vec{x} \cdot \vec{y} > 0$. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P11.3 True or False: If $\vec{x}$ is a vector in $\mathbb{R}^n$ then $\|\vec{x}\| \geq 0$. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P11.4 True or False: For a vector $\vec{x}$ in $\mathbb{R}^n$ we have that $\|\vec{x}\| = 0$ if and only if $\vec{x} = \vec{0}$. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P11.5 True or False: For any vectors $\vec{x}, \vec{y}$ in $\mathbb{R}^n$ we have $\|\vec{x} + \vec{y}\| = \|\vec{x}\| + \|\vec{y}\|$. If true, provide a proof. If false, provide a counterexample and justify why this is a counterexample.

P11.6 Prove Proposition 11.2.

P11.7 For vectors $\vec{x}, \vec{y}$ in $\mathbb{R}^n$ show that $d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x})$.

P11.8 Show that if $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ is a set of nonzero orthogonal vectors in $\mathbb{R}^n$, then $\mathcal{B}$ forms a basis for $\mathbb{R}^n$.

P11.9 Show that if A and B are orthogonal matrices, then the product AB is also orthogonal.

P11.10 Show that if Q is orthogonal, then $\det(Q) = \pm 1$.

P11.11 In this problem, we'll show that the 2×2 orthogonal matrices transform $\mathbb{R}^2$ by either reflection or rotation.

(a) Show that for a vector $\vec{v} \in \mathbb{R}^2$, $\|\vec{v}\| = 1$ if and only if $\vec{v}$ is of the form $\vec{v} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$, where θ is the angle $\vec{v}$ makes with the positive x-axis.

(b) Suppose that Q is orthogonal. Use the previous part to show that

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ or } Q = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

(Hint: note that if $Q = \begin{pmatrix} \vec{v} & \vec{w} \end{pmatrix}$ is orthogonal, then the angle $\vec{w}$ makes with the positive x-axis is $\theta \pm \pi/2$, and then apply trig identities.)

(c) Conclude that if Q is a 2×2 orthogonal matrix, then Q is a rotation matrix when $\det(Q) = 1$ and a reflection matrix when $\det(Q) = -1$.

P11.12 Show that any orthogonal set of nonzero vectors $\{\vec{v}_1, \ldots, \vec{v}_m\}$ in $\mathbb{R}^n$ is linearly independent. That is, if $\vec{v}_i \cdot \vec{v}_j = 0$ for all $i \neq j$, then the set $\{\vec{v}_1, \ldots, \vec{v}_m\}$ is linearly independent.

P11.13 Show that the rank of a matrix A is equal to the number of nonzero singular values of A.

P11.14 Show that a square matrix A is invertible if and only if 0 is not a singular value of A.

P11.15 Show that A and $A^\top$ have the same nonzero singular values.

Appendix A

Proof Writing Guides

A.1. Sets and Set Notation

A.2. Mathematical Statements

Oftentimes, mathematicians try to show whether "something" is true or false. Let's be a bit more precise about these objects of study.

Definition A.1. A STATEMENT is a declarative sentence that is either true or false, but not both.

Example A.2. The sentence "2+2=4" is a true statement, and the sentence "2+2=5" is a false statement.

Example A.3. Note that the sentence "2x+1=5" is **not** a statement, since its truth value depends on x. We could edit this sentence to make it a statement by adding a particular value of x. For example, "If $x = 2$ then $2x + 1 = 5$" is now a statement.

Definition A.4. PROVING a statement means to demonstrate, using sound logical reasoning, that a statement is true. DISPROVING a statement means to demonstrate, using sound logical reasoning, that statement is false.

There are a few synonyms for "statement" that you should be aware of:

(1) PROPOSITION: This is the generic term for "statement";
(2) THEOREM: These are statements that have some significance in a piece of work. We'll often use the word theorem to highlight the importance of a result or to indicate that it requires a more involved proof than other statements in the work.
(3) LEMMA: These are usually "helper statements". That is, they are statements which help you prove larger results, like propositions and theorems.
(4) COROLLARY: These are statements that follow quickly, or without too much work, from a proposition or theorem.

A common statement structure you'll run across takes the form: "if *blah* then *blah*". Let's give a name to this type of statement.

Definition A.5. A CONDITIONAL STATEMENT is any statement of the form "if P then Q", where P and Q are both statements.

To prove a conditional statement "if P then Q", we need to do the following:

(1) Assume the hypothesis P is true,
(2) Use logical/mathematical reasoning to deduce the conclusion Q is true.

Example A.6. Let's prove the following conditional statement:

"If $x = 1$, then $x^2 + x + 1 = 3$.

Proof. Suppose that $x = 1$ (here, we assume the hypothesis is true). Observe that

$$1^2 + 1 + 1 = 1 + 1 + 1 = 3.$$

Hence, $x^2 + x + 1 = 3$ (note that the sentence in the middle used mathematical reasoning to show the conclusion holds). □

The proof structure outlined above has some logical backing we won't get into here. If this is a topic you're interested in, I suggest taking a look at CMU Faculty Clive Newstead's text Infinite Descent.

There's one more type of statement we should be aware of for this course.

Definition A.7. A BICONDITIONAL STATEMENT is any statement of the form

"P if and only if Q"

and is equivalent to the statement

"If P then Q and if Q then P".

To prove a biconditional statement "P if and only if Q", we need to:

(1) Prove the conditional statement "If P then Q" **and**
(2) Prove the conditional statement "If Q then P"

A.3. Set Equality

Recall that a SET is an collection of objects. The objects contained in a set are called its ELEMENTS

Example A.8. $\{1, 2, x\}$ is a set containing the elements 1, 2 and x.

Definition A.9. Let A and B be sets. We say that A is a SUBSET of B if every element of A is also an element of B. We say that sets A and B are EQUAL if they have the same elements.

If we want to prove two sets A and B are equal, we need to show that A and B have exactly the same elements. One way to do this is to show that every element of A is an element of B, and vice versa. That is, to show set equality $A = B$, a standard approach is to prove two set inclusions: $A \subseteq B$ and $B \subseteq A$.

Example A.10. Let $A = \{2k+1 \mid k \text{ is an integer}\}$ and $B = \{2\ell-1 \mid \ell \text{ is an integer}\}$. Let's prove that $A = B$.

Proof. Take any $a \in A$. Then there's an integer k so that $a = 2k + 1$. We have

$$a = 2k + 1 = 2k + 2 - 1 = 2(k + 1) - 1 = 2\ell - 1.$$

Hence, a is an element of B, and so $A \subseteq B$. Conversely, suppose that $b \in B$. Then there exists an integer ℓ so that $b = 2\ell - 1$. We have

$$b = 2\ell - 1 = 2\ell - 2 + 1 = 2(\ell - 1) + 1 = 2k + 1$$

where $k = \ell + 1$. Hence, b is an element of A, and so $B \subseteq A$. This gives $A = B$ as needed. □

A.4. Proof by Contradiction

Recall that a conditional statement is a statement of the form "if P then Q". We learned how to prove a conditional statement directly in Section A.2, by assuming the hypothesis P is true, and using some logical arguments to conclude that the conclusion Q must also be true.

Sometimes, proving a statement directly is out of reach. Instead, we have the following crafty strategy. Note that this works to prove any type of statement (not just conditional statements).

To prove statement X is true, we can do the following.

(1) Assume that X is **false**.

(2) Use some logical reasoning to show that this implies a contradiction (that is, some nonsense that certainly isn't true, like $2 = 4$ or π is an integer, etc).

(3) Conclude that statement X must have been true.

Hopefully, it feels intuitive that the only way for a statement to imply something false is if the statement itself was false (note that the statement "X is false" being false means that X is true). But if you're interested in why proof by contradiction is valid using formal logic, I would look at Section 3.3 of Sundstrum's text.

Let's look at a quick example of proof by contradiction.

Proposition A.11. The sum of a rational number and an irrational number is irrational.

Proof. Suppose that x is rational and y is irrational. Then we can write $x = a/b$ for integers a and b with $b \neq 0$. For a contradiction, suppose that $x + y$ is rational. Then, there exist integers c and $d \neq 0$ so that $x + y = c/d$. We get

$$\frac{a}{b} + y = \frac{c}{d}$$

$$\Rightarrow y = \frac{c}{d} - \frac{a}{b} = \frac{cb - ad}{bd},$$

which implies that y is rational. Since a real number cannot be both rational and irrational, we have a contradiction. So, it must be the case that $x + y$ is irrational. $\square$

Appendix B

Activity Solutions

B.1. Chapter 1 Activity Solutions

Activity 1.1. The following examples have been adapted from Chapter 2 of [**2**], where interested students can find further applications.

Allocation of Resources. A biologist has three strains of bacteria (I, II, and III), which will be placed in a test tube to feed on three different food sources (A, B, and C). The biologist places 230 grams of Food A, 80 grams of Food B, and 150 grams of Food C into the test tube. Given that each bacterium consumes the amount of food indicated in the table below per day, how many bacteria of each strain can coexist in the test tube and consume all of the food?

	Strain I	Strain II	Strain III
Food A	2 g	2 g	4 g
Food B	1 g	2 g	0 g
Food C	1 g	3 g	1 g

Solution. Suppose that we have x amount of Strain I, y amount of Strain II, and z amount of Strain III. Then, we need x, y, and z to satisfy each of the equations below

$$\begin{cases} 2x + 2y + 4z = 230 \\ x + 2y = 80 \\ x + 3y + z = 150 \end{cases}$$

Using some math (at this moment, we only have ad-hoc methods available), we obtain $\boxed{x = 10, y = 35, z = 35}$.

Input-Output Analysis. Three neighbors, each with a vegetable garden, agree to share their produce. One will grow beans, one will grow lettuce, and one will grow tomatoes. The following table shows what percentage of each crop each of the neighbors will receive.

	bean grower	lettuce grower	tomato grower
% beans shared	25%	50%	25%
% lettuce shared	40%	30%	30%
% tomato shared	10%	60%	30%

Given this arrangement, explain how the neighbors are valuing each item of produce so that they all break even in this exchange.

Solution. Suppose that the neighbors value beans at b dollars per crop, lettuce at ℓ dollars per crop, and tomatoes at t dollars per crop. Then, in the exchange above, for each crop the bean grower earns

$$0.25b + 0.4\ell + 0.1t$$

the lettuce grower earns

$$0.5b + 0.3\ell + 0.6t$$

and the tomato grower earns

$$0.25b + 0.3\ell + 0.3t.$$

Now, for the bean grow to break even, what they earn (listed above) should be equal in value to what they produce (b). That is, for the bean grower to break even, we need

$$0.25b + 0.4\ell + 0.1t = b.$$

Similarly, for the lettuce grower to break even we need

$$0.5b + 0.3\ell + 0.6t = \ell,$$

and for the tomato grower to break even we need

$$0.25b + 0.3\ell + 0.3t = t.$$

Hence, we want to find values for b, ℓ, and t so that all of the equations are satisfied simultaneously

$$\begin{cases} 0.25b + 0.4\ell + 0.1t = b \\ 0.5b + 0.3\ell + 0.6t = \ell \\ 0.25b + 0.3\ell + 0.3t = t. \end{cases}$$

With some math, we obtain that $\boxed{b \approx 0.95t \text{ and } \ell \approx 1.5t}$. That is, the neighbors value beans slightly less than tomatoes, and lettuce 1.5 times as much as tomatoes.

Activity 1.2 Find all solutions to the the system of linear equations that have the following augmented matrices

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

Solution. This augmented matrix represents the system

$$\begin{cases} x + y + z = 1 \\ \qquad y - z = 1 \\ \qquad\quad 2z = 2 \end{cases}$$

From the last equation we get

$$2z = 2 \Rightarrow z = 1.$$

Substituting this into the second equation gives

$$y - 1 = 1 \Rightarrow y = 2.$$

Finally, substitution $z = 1$ and $y = 2$ into the first equation gives

$$x + 2 + 1 = 1 \Rightarrow x = -2.$$

Hence, this system has one solution, given by $\boxed{(-2, 2, 1)}$.

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

Solution. This augmented matrix represents the system

$$\begin{cases} 2y - z = 1 \\ 0 = 1 \end{cases}$$

But, there does not exist a triple of real numbers (x, y, z) so that $0 = 1$. Hence, this system has $\boxed{\text{no solutions}}$.

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

Solution. This augmented matrix represents the system

$$\begin{cases} 2x + 2y + z = 2 \\ 3z = 6. \end{cases}$$

From the last equation we get

$$3z = 6 \Rightarrow z = 2.$$

Substituting this into the first equation gives

$$2x + 2y + 2 = 2 \Rightarrow x = 1 - y.$$

So, the solutions to the equation must take the form $\boxed{(1 - y, y, 2) \text{ where } y \text{ is any real number}}$.

Activity 1.3. Determine which of the following matrices are in row echelon form, reduced row echelon form, or neither.

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 1 \\ 3 & 4 & -1/2 \end{pmatrix}, \; B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; C = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{pmatrix},$$

$$D = \begin{pmatrix} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; E = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; F = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

Solution. The matrices B, C, D and F are in row echelon form. The matrices B and F are in reduced row echelon form.

Activity 1.4. Give examples of 3×4 matrices that satisfy the following conditions. You should provide a new matrix for each part.

(1) In row echelon form, but not reduced row echelon form.

Sample Solution.
$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 4 & 5 \\ 0 & 0 & 1 & 2 \end{pmatrix}.$$

(2) In reduced row echelon form.

Sample Solution.
$$\begin{pmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 2 \end{pmatrix}.$$

(3) In reduced row echelon form and has exactly one row of zeros.

Sample Solution.
$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

(4) In reduced row echelon form, has exactly one pivot, and has at least one entry other than 1 or 0.

Sample Solution.
$$\begin{pmatrix} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

B.2. Chapter 2 Activity Solutions

Activity 2.1. Suppose that someone gave you the following directions: from your starting point, first walk two blocks west and one block north. After that, walk one block east and then three blocks north. Find the standard coordinate representation of your total displacement.

Solution. Our total displacement vector is given by

$$\begin{pmatrix} -2+1 \\ 1+3 \end{pmatrix} = \boxed{\begin{pmatrix} -1 \\ 4 \end{pmatrix}}.$$

That is, in total we've walked 1 block west and 4 blocks north.

Activity 2.2 (The Magic Carpet Ride, Scenario 1). You are the recipient of a new scholarship at the University of Toronto. As part of your scholarship, you have been gifted two forms of transportation to help you navigate the city: a hover board and a magic carpet. Upon reading the instructions, you find that your two modes of transportation have restrictions on how they operate:

- The hoverboard can only move along the vector $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$

  By this we mean that if the hover board traveled "forward" for one hour, it would move along a "diagonal" path that would result in a displacement of 3 miles East and 1 mile North of its starting location.

- The magic carpet can only move along the vector $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

  By this we mean that if the magic carpet traveled "backward" for one hour, it would move along a "diagonal" path that would result in a displacement of 1 mile West and 2 miles South of its starting location.

Scenario One: Your friend Cramer suggests that you should go visit his friend Gauss outside of Toronto. Cramer tells you that Gauss lives in a cabin that's 107 miles East and 64 miles North of your home here in Toronto.

(1) Is it possible to reach Gauss's cabin using only the hoverboard? If so, how? If not, why not?

Solution. Suppose that we spend x hours on the hoverboard. Then, to reach Gauss' cabin observe that the following vector equation would need to be satisfied

$$x\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 107 \\ 64 \end{pmatrix}.$$

But this implies we would need a solution to the system of linear equations

$$\begin{cases} 3x = 107 \\ \ \ x = 64. \end{cases}$$

Since this system does not have a solution, it is $\boxed{\text{not possible}}$ to reach Gauss' cabin using only the hoverboard.

(2) Is it possible to reach Gauss's cabin using only the magic carpet? If so how? If not, why not?

Solution. Suppose that we spend y hours on the magic carpet. Then, to reach Gauss' cabin observe that the following vector equation would need to be satisfied

$$y\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}107\\64\end{pmatrix}.$$

But this implies we would need a solution to the system of linear equations

$$\begin{cases} \ \ y = 107 \\ 2y = 64. \end{cases}$$

Since this system does not have a solution, it is $\boxed{\text{not possible}}$ to reach Gauss' cabin using only the hoverboard.

(3) Is it possible to reach Gauss's cabin using both the hoverboard and magic carpet? If so how? If not, why not?

Solution. Suppose that we spend x hours on the hoverboard and y hours on the magic carpet Then, to reach Gauss' cabin observe that the following vector equation would need to be satisfied

$$x\begin{pmatrix}3\\1\end{pmatrix} + y\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}107\\64\end{pmatrix}.$$

This has the same solution set as the system of linear equations

$$\begin{cases} 3x + y = 107 \\ \ \ x + 2y = 64, \end{cases}$$

which has corresponding augmented matrix

$$\left(\begin{array}{cc|c} 3 & 1 & 107 \\ 1 & 2 & 64 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 30 \\ 0 & 1 & 17 \end{array}\right).$$

Hence, it is possible to reach Gauss' cabin, by spending $\boxed{\text{30 hours}}$ on the hoverboard and $\boxed{\text{17 hours}}$ on the magic carpet.

Activity 2.3 (The Magic Carpet Ride, Scenario 2 Recall that you have two modes of transportation: a hoverboard and a magic carpet.

- The hoverboard can only move along the vector $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$
- The magic carpet can only move along the vector $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

Scenario Two: It turns out that Gauss actually does not like visitors at his cabin. After Gauss learns that Cramer shared the location of his cabin, he starts looking for a new place to live so that you don't bother him again.

(1) Is there anywhere Gauss can move so that you cannot reach him using your hoverboard and magic carpet? Describe the places that you can reach using a combination of the hover board and the magic carpet and those you cannot.

Solution. Suppose that Gauss relocated to coordinates (g_1, g_2). As before, we want to solve the vector equation

$$x\begin{pmatrix} 3 \\ 1 \end{pmatrix} + y\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$$

which has the same solution set as the system with augmented matrix

$$\left(\begin{array}{cc|c} 3 & 1 & g_1 \\ 1 & 2 & g_2 \end{array}\right).$$

After row reducing, we find that this matrix has reduced row echelon form

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

where $*$ denotes unknown real numbers. Note that no matter what values g_1, g_2 take, this matrix will never have a pivot in the last column. So, by Rouché-Capelli, the system always has a solution. So, $\boxed{\text{no}}$, there is nowhere Gauss can move where he cannot be reached.

(2) Gauss also starts looking into purchasing a hover cabin, so that you can not reach him by ground transport. Unfortunately for him, your modes of transportation can also fly! Reading the instructions more closely, we find

- In flying mode, your hoverboard can move forward and backward along the vector $\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}$
- In flying mode, your magic carpet can move forward and backward along the vector $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

Is there a location for Gauss to hide in his hover cabin so that you can't reach him using your hoverboard and magic carpet in flying mode?

Solution. Setting up similarly before, we'd like to know whether there are solutions to the vector equation

$$x\begin{pmatrix}3\\1\\1\end{pmatrix} + y\begin{pmatrix}1\\2\\3\end{pmatrix} = \begin{pmatrix}g_1\\g_2\\g_3\end{pmatrix}$$

which has the same solution set as the system of linear equations with augmented matrix

$$\left(\begin{array}{cc|c}3 & 1 & g_1\\1 & 2 & g_2\\1 & 3 & g_3\end{array}\right).$$

After row reducing, we find this matrix has a row of the form

$$\left(\begin{array}{cc|c}0 & 0 & g_1 - 8g_2 + 5g_3\end{array}\right),$$

and so if Gauss relocates his cabin to the coordinates $(8, 1, 1)$, for example, this matrix would have a pivot in the last column, and hence (by Rouché-Capelli) our system would not have a solution. So, $\boxed{\text{yes}}$ it is possible for Gauss to relocate so that we cannot reach him. For example, he could relocate to the coordinates $(8, 1, 1)$.

Activity 2.4. Consider the vectors

$$\vec{v}_1 = \begin{pmatrix}1\\1\\1\end{pmatrix}, \vec{v}_2 = \begin{pmatrix}6\\3\\8\end{pmatrix}, \vec{v}_3 = \begin{pmatrix}4\\1\\6\end{pmatrix}$$

in $\mathbb{R}^3$. Is $\operatorname{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ equal to a line, a plane, or all of $\mathbb{R}^3$?

Solution. Observe that we have

$$\vec{v}_3 = \vec{v}_2 - 2\vec{v}_1$$

and so it can be seen that

$$\operatorname{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3) = \operatorname{Span}(\vec{v}_1, \vec{v}_2).$$

Since $\vec{v}_1, \vec{v}_2$ are distinct and non-parallel, this span forms a $\boxed{\text{plane}}$ in $\mathbb{R}^3$.

Activity 2.5. Show that our two definitions of linear dependence agree. That is,

(1) Explain how the geometric definition of linear dependence implies the algebraic one.

Solution. Suppose that $\vec{v}_1, \ldots, \vec{v}_n$ are linearly dependent in the geometric sense. Without loss of generality, we may assume that

$$\vec{v}_1 \in \operatorname{Span}(\vec{v}_2, \ldots, \vec{v}_n).$$

Hence, there are constants $c_2, \ldots, c_n \in \mathbb{R}$ so that

$$\vec{v}_1 = c_2\vec{v}_2 + \cdots + c_n\vec{v}_n$$

$$\Rightarrow \vec{0} = -\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n.$$

Hence, $(-1, c_2, \ldots, c_n)$ is a solution to the vector equation

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{0}.$$

Since $-1 \neq 0$ this solution is nontrivial, and so our vectors are linearly independent in the algebraic sense. $\square$

(2) Explain how the algebraic definition of linear dependence implies the geometric one.

Solution. Suppose that $\vec{v}_1, \ldots, \vec{v}_n$ are linearly dependent in the algebraic sense. Then, there is a nontrivial solution $(c_1, \ldots, c_n)$ to the vector equation

$$x_1\vec{v}_1 + \cdots + x_n\vec{v}_n = \vec{0}.$$

Since our solution is nontrivial, we must have $c_i \neq 0$ for some $i \in \{1, \ldots, n\}$. By relabeling, we may assume that $c_1 \neq 0$. Hence, we have

$$\vec{v}_1 = \frac{c_2}{-c_1}\vec{v}_2 + \cdots + \frac{c_n}{-c_2},$$

and so $\vec{v}_1 \in \mathrm{Span}(\vec{v}_2, \ldots, \vec{v}_n)$, and so our vectors are linearly dependent in the geometric sense, as needed. $\square$

B.3. Chapter 3 Activity Solutions

Activity 3.1. Consider the following subsets of $\mathbb{R}^3$, which are graphed in desmos. Use geometric reasoning to determine whether the sets could be equal to a span.

(1) link to desmos graph of $\mathcal{S}$

(2) link to desmos graph of $\mathcal{T}$

(3) link to desmos graph of $\mathcal{R}$

Solution.

(1) The set $\mathcal{S}$ cannot be equal to a span. Think of a vector originating from the origin and ending at the point $(1, 0, 0)$. This vector is in $\mathcal{S}$ but if we scale it by any scalar greater than 2, the resulting vector will not be in $\mathcal{S}$. If a vector is in a span, then any scalar multiple of it should also be in the span, (since the linear combination would just have the same coefficients but scaled by the scalar).

(2) The set $\mathcal{T}$ can be equal to a span. It is the xy-plane so every point on it can be expressed as a linear combination of the vectors $\begin{pmatrix}1\\0\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\\0\end{pmatrix}$.

(3) The set $\mathcal{R}$ cannot be equal to a span. If we take two random nonzero vectors on each of these planes, and we try to add them, the resulting vector will not be in $\mathcal{R}$. But if $\mathcal{R}$ is a span, then each point that we chose (say $\vec{u}$ and $\vec{v}$) would be a linear combination of some fixed vectors. Adding the two vectors should still give us a linear combination of the fixed vectors, since we would simply be adding the coefficients of the linear combinations of the original vectors ($\vec{u}$ and $\vec{v}$) to get the coefficients of the linear combination for the resulting vector $(\vec{u} + \vec{v})$.

Activity 3.2. Consider the following subset of $\mathbb{R}^3$

$$V = \left\{ \begin{pmatrix} x-y \\ x+y \\ y \end{pmatrix} : x, y \in \mathbb{R} \right\}.$$

(1) Show that V is a vector space by using Definition 3.1.

(2) Show that V is a vector space by finding a generating set for V.

Solution.

(1) First note that V is non-empty, since the vector $\begin{pmatrix}0\\2\\1\end{pmatrix} \in V$ (where we take $x = y = 1$). Now let $\vec{u}, \vec{v} \in V$ and let $c \in \mathbb{R}$ be a scalar. Then for some $u_1, u_2, v_1, v_2 \in \mathbb{R}$, we have

$$\vec{u} = \begin{pmatrix} u_1 - u_2 \\ u_1 + u_2 \\ u_2 \end{pmatrix}, \vec{v} = \begin{pmatrix} v_1 - v_2 \\ v_1 + v_2 \\ v_2 \end{pmatrix}.$$

Now

$$\begin{aligned}
\vec{u}+\vec{v} &= \begin{pmatrix} u_1-u_2 \\ u_1+u_2 \\ u_2 \end{pmatrix} + \begin{pmatrix} v_1-v_2 \\ v_1+v_2 \\ v_2 \end{pmatrix} \\
&= \begin{pmatrix} u_1-u_2+v_1-v_2 \\ u_1+u_2+v_1+v_2 \\ u_2+v_2 \end{pmatrix} \\
&= \begin{pmatrix} (u_1+v_1)-(u_2+v_2) \\ (u_1+v_1)+(u_2+v_2) \\ (u_2+v_2) \end{pmatrix} \\
&= \begin{pmatrix} x-y \\ x+y \\ y \end{pmatrix}, \text{ where } x=u_1+v_1, y=u_2+v_2.
\end{aligned}$$

And so $\vec{u}+\vec{v} \in V$ and V is closed under vector addition. Now to check closure under scalar multiplication,

$$\begin{aligned}
c\vec{u} &= c\begin{pmatrix} u_1-u_2 \\ u_1+u_2 \\ u_2 \end{pmatrix} \\
&= \begin{pmatrix} c(u_1-u_2) \\ c(u_1+u_2) \\ cu_2 \end{pmatrix} \\
&= \begin{pmatrix} cu_1-cu_2 \\ cu_1+cu_2 \\ cu_2 \end{pmatrix} \\
&= \begin{pmatrix} x-y \\ x+y \\ y \end{pmatrix}, \text{ where } x=cu_1, y=uc_2.
\end{aligned}$$

It follows that V is closed under scalar multiplication. Since V satisfies all the conditions in Definition 3.1, V is a vector space.

(2) Note that any vector in V has the form

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

where $x, y \in \mathbb{R}$. But this means that $V = \text{Span}\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}\right)$. Since V has a generating set, it is a vector space by Theorem 3.4.

Activity 3.3. Determine which of the following sets are bases for $\mathbb{R}^3$

$$\mathcal{B}_1 = \left\{ \begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}, \begin{pmatrix}1\\2\\2\end{pmatrix} \right\}$$

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

$$\mathcal{B}_3 = \left\{ \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix} \right\}$$

$$\mathcal{B}_4 = \left\{ \begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix} \right\}$$

Solution. Note that

$$\begin{pmatrix}1\\2\\2\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix} + \begin{pmatrix}0\\1\\1\end{pmatrix},$$

and so the vectors in $\mathcal{B}_1$ are linearly dependent, and $\mathcal{B}_1$ is not a basis for $\mathbb{R}^3$.

Since the vector

$$\begin{pmatrix}0\\0\\1\end{pmatrix} \notin \operatorname{Span}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix} \right),$$

it follows that $\mathcal{B}_2$ is not a generating set for $\mathbb{R}^3$, and therefore is not a basis for $\mathbb{R}^3$.

The set $\mathcal{B}_3$ is a basis for $\mathbb{R}^3$, since its vectors are linearly independent and are a generating set for $\mathbb{R}^3$.

The set $\mathcal{B}_4$ is a basis for $\mathbb{R}^3$. First we show that the vectors are linearly independent. Looking at the augmented matrix

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

we see that it has RREF

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

And so the vector equation

$$a_1 \begin{pmatrix}1\\0\\1\end{pmatrix} + a_2 \begin{pmatrix}1\\1\\0\end{pmatrix} + a_3 \begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}$$

has only the trivial solution. Therefore, the vectors in $\mathcal{B}_4$ are linearly independent. Now to show that they are a generating set for $\mathbb{R}^3$. Consider an arbitrary vector

$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in R^3$, if the vector equation

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

has a solution, then $\mathcal{B}_4$ is a generating set. The vector equation has the following corresponding augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 0 & 1 & 0 & y \\ 1 & 0 & 1 & z \end{array}\right),$$

which has RREF

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & x-y \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & z-x+y \end{array}\right).$$

And so the vector equation has solution $a = x - y, b = y, c = z - x + y$. Since a solution exists for any arbitrary vector in $\mathbb{R}^3$ it follows that $\mathcal{B}_4$ is a generating set for $\mathbb{R}^3$.

Activity 3.4. Show that $\dim \mathbb{R}^n = n$.

Solution. We show this by finding a basis for $\mathbb{R}^n$. Consider the vectors

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \vec{e}_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$

where the vector $\vec{e}_i$ is the vector whose i^{th} entry is 1, and all other entries are 0. These vectors are linearly independent. If we look at the vector equation

$$a_1\vec{e}_1 + \cdots + a_n\vec{e}_n = \vec{0}$$

we can see that it only has the trivial solution. These vectors are also a generating set for $\mathbb{R}^n$, since any arbitrary vector in $\mathbb{R}^n$ can be expressed as

$$\begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} = x_1\vec{e}_1 + \cdots + x_i\vec{e}_i + \cdots + x_n\vec{e}_n.$$

Therefore, the vectors $\{\vec{e}_1, \ldots, \vec{e}_n\}$ form a basis for $\mathbb{R}^n$ and so $\dim \mathbb{R}^n = n$.

Activity 3.5. Let $V = \mathrm{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4)$ be a vector subspace of $\mathbb{R}^n$ and

$$A = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 & \vec{v}_4 \end{pmatrix}.$$

Suppose that $\mathrm{rref}(A)$ has a pivot in columns $1, 3$ and 4 but does not have a pivot in column 2.

(1) Show that $\vec{v}_2 \in \mathrm{Span}(\vec{v}_1, \vec{v}_3, \vec{v}_4)$

(2) Show that $\{\vec{v}_1, \vec{v}_3, \vec{v}_4\}$ is a linearly independent set

(3) Conclude that $\dim V = 3$.

Solution.

(1) Consider the vector equation

$$a_1\vec{v}_1 + a_2\vec{v}_2 + a_3\vec{v}_3 + a_4\vec{v}_4 = \vec{0}.$$

Since there is not pivot in the second column, of $\mathrm{rref}(A)$, we know that the variable a_2 is free. So we can set a_2 to any scalar and get a solution to the vector equation above. Setting $a_1 = -1$, we get

$$a_1\vec{v}_1 - \vec{v}_2 + a_3\vec{v}_3 + a_4\vec{v}_4 = \vec{0}$$

and rearranging, we get

$$\vec{v}_2 = a_1\vec{v}_1 + a_3\vec{v}_3 + a_4\vec{v}_4 \in \mathrm{Span}(\vec{v}_1, \vec{v}_3, \vec{v}_4.$$

(2) Since we found that $\vec{v}_2$ is redundant, we may remove it from the matrix A and not change the span of the columns of the matrix. Removing $\vec{v}_2$ from A gives us a matrix A' with three pivot columns. Consider the vector equation

$$a_1\vec{v}_1 + a_3\vec{v}_3 + a_4\vec{v}_4 = \vec{0}$$

Since $\mathrm{rref}(A')$ has a pivot in every column, it follows that the vector equation above has only the trivial solution. Therefore, the set $\{\vec{v}_1, \vec{v}_3, \vec{v}_4\}$ is a linearly independent set.

(3) We have shown that $\{\vec{v}_1, \vec{v}_3, \vec{v}_4\}$ is a linearly independent set. We have also shown that $\mathrm{Span}(\vec{v}_1, \vec{v}_3, \vec{v}_4) = \mathrm{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4) = V$. This means that $\{\vec{v}_1, \vec{v}_3, \vec{v}_4\}$ is a linearly independent generating set for V, and so is a basis for V. Since a basis for V contains 3 vectors, it follows that $\dim V = 3$.

B.4. Chapter 4 Activity Solutions

Activity 4.1. Let
$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}.$$

(1) Find a vector $\vec{v} \in \mathrm{Col}(A)$.

(2) Find a vector $\vec{w} \in \mathrm{Nul}(A)$.

Solution.

(1) Any linear combination of the column vectors of A is a vector in $\mathrm{Col}(A)$. For example,
$$\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} + \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \\ 11 \end{pmatrix} \in \mathrm{Col}(A).$$

(2) Any vector in the solution space for $A\vec{x} = \vec{0}$ is a vector in $\mathrm{Nul}(A)$. Note that
$$\mathrm{rref}(A) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix},$$
which has a pivot in every column. So $A\vec{x} = \vec{0}$ has a unique solution, which is the trivial solution, and $\mathrm{Nul}(A) = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$.

Activity 4.2. Show that $\mathrm{Col}(A)$ and $\mathrm{Nul}(A)$ are in fact vector spaces.

Solution. Let A be an $m \times n$ matrix with
$$A = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}.$$
Then $\mathrm{Col}(A) := \mathrm{Span}(\vec{v}_1, \ldots, \vec{v}_n)$ is a vector space by Theorem 3.4. Next, consider $\mathrm{Nul}(A) := \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}\}$. Note that $\vec{0} \in \mathrm{Nul}(A)$, because $A\vec{0} = \vec{0}$, and so $\mathrm{Nul}(A)$ is nonempty. Now suppose $\vec{x}_1, \vec{x}_2 \in \mathrm{Nul}(A)$. Then $A\vec{x}_1 = A\vec{x}_2 = \vec{0}$ and so
$$A(\vec{x}_1 + \vec{x}_2) = A\vec{x}_1 + A\vec{x}_2 = \vec{0} + \vec{0} = \vec{0}.$$
So $\vec{x}_1 + \vec{x}_2 \in \mathrm{Nul}(A)$ and $\mathrm{Nul}(A)$ is closed under vector addition. Now let c be a scalar, then
$$A(c\vec{x}_1) = c(A\vec{x}_1) = c\vec{0} = \vec{0},$$
and so $c\vec{x}_1 \in \mathrm{Nul}(A)$ and $\mathrm{Nul}(A)$ is closed under scalar multiplication. Since $\mathrm{Nul}(A)$ is a nonempty subset of $\mathbb{R}^n$ that is closed under vector addition and scalar multiplication, it follows that $\mathrm{Nul}(A)$ is a vector space.

Activity 4.4. Find the rank and nullity of the following matrices.

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

(2) $B = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 3 & 2 & -1 \\ 0 & -1 & -1 & 1 \\ 0 & 1 & 1 & -1 \end{pmatrix}$

Solution.

(1) Since A is in RREF, and it has pivots in column 1 and 3, by Theorem 3.11, these two columns form a basis for $\operatorname{Col}(A)$, which is the span of the column vectors of A. Thus $\operatorname{rank}(A) = 2$. Next, note that

$$\vec{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \operatorname{Nul}(A) \text{ if and only if } A\vec{x} = \begin{pmatrix} x-y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

So we need $x - y = 0$ and $z = 0$. Letting y be free, we see that $\vec{x} \in \operatorname{Nul}(A)$ if $\vec{x} = \begin{pmatrix} y \\ y \\ 0 \end{pmatrix}$. This gives

$$\begin{aligned} \operatorname{Nul}(A) &= \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 : A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \vec{0} \right\} \\ &= \left\{ \begin{pmatrix} y \\ y \\ 0 \end{pmatrix} : y \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} y : y \in \mathbb{R} \right\} \\ &= \operatorname{Span} \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\}. \end{aligned}$$

Since $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ on its own is linearly independent, it forms a basis for $\operatorname{Nul}(A)$. Therefore, $\operatorname{nullity}(A) = 1$.

(2) Note that B has reduced row echelon form

$$B' = \begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since $\operatorname{rref}(B)$ only has pivots in columns 1 and 2, by Theorem 3.11, columns 1 and 2 of B form a basis for the column space of B. Therefore, $\operatorname{rank}(B) = 2$.

Next, note that

$$\vec{x} = \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \in \operatorname{Nul}(B) \text{ if and only if } B'\vec{x} = \begin{pmatrix} x - z + 2w \\ y + z - w \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Letting z and w be free, we get solutions $\vec{x} = \begin{pmatrix} z-2w \\ -z+w \\ z \\ w \end{pmatrix}$. So we have

$$\begin{aligned}
\mathrm{Nul}(B) &= \left\{ \begin{pmatrix} z-2w \\ -z+w \\ z \\ w \end{pmatrix} : z, w \in \mathbb{R} \right\} \\
&= \left\{ \begin{pmatrix} z \\ -z \\ z \\ 0 \end{pmatrix} + \begin{pmatrix} -2w \\ w \\ 0 \\ w \end{pmatrix} : z, w \in \mathbb{R} \right\} \\
&= \left\{ \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} z + \begin{pmatrix} -2 \\ 1 \\ 0 \\ 1 \end{pmatrix} w : z, w \in \mathbb{R} \right\} \\
&= \mathrm{Span}\left(\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \\ 1 \end{pmatrix} \right)
\end{aligned}$$

Therefore, $\mathrm{nullity}(B) = \dim(\mathrm{Nul}(B)) = 2$.

Activity ??. Let A be a 3×3 matrix.

(1) Suppose that A has exactly two pivot columns, which are located in columns 1 and 3. Show that $\mathrm{rank}(A) = 2$ and $\mathrm{nullity}(A) = 1$.

(2) Suppose that A has exactly one pivot column, which is located in column 1. Show that $\mathrm{rank}(A) = 1$ and $\mathrm{nullity}(A) = 2$.

Solution.

(1) Since the first and third columns of A are pivot columns, by Theorem 3.11 we know that column 1 and column 3 form a basis for $\mathrm{Col}(A)$ and so

$$\mathrm{rank}(A) = \dim(\mathrm{Col}(A)) = 2.$$

Let $A = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{pmatrix}$. Then

$$\vec{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathrm{Nul}(A) \text{ if and only if } \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \vec{0}.$$

So we must have $x\vec{v}_1 + y\vec{v}_2 + z\vec{v}_3 = \vec{0}$. Since column 2 is not a pivot column, we know that we can take y to be a free variable in that last equation. So that $x = ay$ and $z = by$ for some $a, b \in R$. Then

$$\begin{aligned}\mathrm{Nul}(A) &= \left\{ \begin{pmatrix} ay \\ y \\ by \end{pmatrix} : y \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} a \\ 1 \\ b \end{pmatrix} y : y \in \mathbb{R} \right\} \\ &= \mathrm{Span}\left(\begin{pmatrix} a \\ 1 \\ b \end{pmatrix} \right).\end{aligned}$$

This means that $\mathrm{nullity}(A) = \dim(\mathrm{Nul}(A)) = 1$.

(2) Since the first column of A is the only pivot column, by Theorem 3.11, column 1 forms a basis for $\mathrm{Col}(A)$, and $\mathrm{rank}(A) = \dim(\mathrm{Col}(A)) = 1$. Next, let

$$A = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{pmatrix}.$$

Then $\vec{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathrm{Nul}(A)$ if and only if $\begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \vec{0}$. So we must have

$$x\vec{v}_1 + y\vec{v}_2 + z\vec{v}_3 = \vec{0}.$$

Since column 2, and column 3 are not pivot columns, we know that we can take y and z to be free variables in that last equation. So that $x = ay + bz$ for some $a, b \in R$. Then

$$\begin{aligned}\mathrm{Nul}(A) &= \left\{ \begin{pmatrix} ay + bz \\ y \\ z \end{pmatrix} : y, z \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} ay \\ y \\ 0 \end{pmatrix} + \begin{pmatrix} bz \\ 0 \\ z \end{pmatrix} : y, z \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} a \\ 1 \\ 0 \end{pmatrix} y + \begin{pmatrix} b \\ 0 \\ 1 \end{pmatrix} z : y \in \mathbb{R} \right\} \\ &= \mathrm{Span}\left(\begin{pmatrix} a \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} b \\ 0 \\ 1 \end{pmatrix} \right).\end{aligned}$$

Note that the vectors $\begin{pmatrix} a \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} b \\ 0 \\ 1 \end{pmatrix}$ are not multiples of each other, and therefore, are linearly independent. So they form a basis for the null space of B. This means that $\mathrm{nullity}(B) = \dim(\mathrm{Nul}(B)) = 2$.

Activity 4.5. Determine whether the solution set for each of the following systems is empty, a point, a line, or a plane in $\mathbb{R}^3$.

(1) System 1

$$\begin{cases} x + 2y + 4z = 1 \\ x + y - z = 2 \\ y + 5z = -1 \end{cases}$$

(2) System 2

$$\begin{cases} x + 2y + 2z = 5 \\ x + y + z = 0 \\ 3x + 3z = 1 \end{cases}$$

(3) System 3

$$\begin{cases} x + 2y + 4z = a \\ x + y - z = b \\ y + 5z = c \end{cases}$$

(4) System 4

$$\begin{cases} x + 2y + 2z = a \\ x + y + z = b \\ 3x + 3z = c \end{cases}$$

Note that a, b, and c denote unknown real numbers.

Solution.

(1) System 1: Let's take a look at the corresponding augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 2 & 4 & 1 \\ 1 & 1 & -1 & 2 \\ 0 & 1 & 5 & -1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & -6 & 3 \\ 0 & 1 & 5 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

There is no pivot in the last column of the RREF of the augmented matrix, so the system is consistent. The coefficient matrix has one free variable, which means that $\mathrm{Nul}(A)$, where A is the coefficient matrix is 1-dimensional. So the solution set will be 1-dimensional, in other words, it will be a line.

(2) System 2: Let's take a look at the corresponding augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 2 & 2 & 5 \\ 1 & 1 & 1 & 0 \\ 3 & 0 & 3 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & -5 \\ 0 & 1 & 0 & -\frac{1}{3} \\ 0 & 0 & 1 & \frac{16}{3} \end{array}\right).$$

There is no pivot in the last column of the RREF of the augmented matrix, so the system is consistent. The coefficient matrix has no free variables, which means that $\mathrm{Nul}(A)$, where A is the coefficient matrix is 0-dimensional. So the solution set will be 0-dimensional, in other words, it will be a point.

(3) System 3: Let's take a look at the corresponding augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 2 & 4 & a \\ 1 & 1 & -1 & b \\ 0 & 1 & 5 & c \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & -6 & 2b - a \\ 0 & 1 & 5 & a - b \\ 0 & 0 & 0 & -a + b + c \end{array}\right).$$

There is possibly a pivot in the last column of the RREF of the augmented matrix, so the system may be inconsistent. If $-a + b + c = 0$ then the system is consistent. In that case, the coefficient matrix has one free variable, which means that $\mathrm{Nul}(A)$, where A is the coefficient matrix is 1-dimensional. So the solution set will be 1-dimensional, in other words, it will be a line.

(4) System 4: Let's take a look at the corresponding augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 2 & 2 & a \\ 1 & 1 & 1 & b \\ 3 & 0 & 3 & c \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2b - a \\ 0 & 1 & 0 & b - \frac{c}{3} \\ 0 & 0 & 1 & a - 2b + \frac{c}{3} \end{array}\right).$$

There is no pivot in the last column of the RREF of the augmented matrix, so the system is consistent. The coefficient matrix has no free variables, which means that $\mathrm{Nul}(A)$, where A is the coefficient matrix is 0-dimensional. So the solution set will be 0-dimensional, in other words, it will be a point.

B.5. Chapter 5 Activity Solutions

Activity 5.1. Determine which of the following are linear transformations

(1) F: $\mathbb{R}^2 \to \mathbb{R}^2$ defined by $F(\vec{x}) = 2\vec{x}$

(2) $G : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$G\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x^2 \\ y^2 \end{pmatrix}$$

(3) $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$T\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x+1 \\ y-1 \end{pmatrix}.$$

Solution.

(1) We claim that F is linear. To see this, take any $\vec{x}, \vec{y} \in \mathbb{R}^2$. Then,

$$\begin{aligned} F(\vec{x}+\vec{y}) &= 2(\vec{x}+\vec{y}) \\ &= 2\vec{x} + 2\vec{y} \\ &= F(\vec{x}) + F(\vec{y}), \end{aligned}$$

and for any scalar $c \in \mathbb{R}$ we have

$$\begin{aligned} F(c\vec{x}) &= 2(c\vec{x}) \\ &= c(2\vec{x}) \\ &= cF(\vec{x}), \end{aligned}$$

as needed.

(2) This function is not linear. For example, we have

$$G\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = G\left(\begin{pmatrix} 2 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 4 \\ 0 \end{pmatrix},$$

but

$$G\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + G\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix},$$

and so

$$G\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) \neq G\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + G\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right),$$

which shows that G is not linear.

(3) This function is not linear. One way to see this is to note that $T(\vec{0}) = T(0\vec{v}) \neq 0T(\vec{v}) = \vec{0}$, where $\vec{v} \in \mathbb{R}^3$. In fact, $T(\vec{0}) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \neq \vec{0}$.

Activity 5.2. Suppose that $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation. Recall that a line in $\mathbb{R}^2$ can be described by the set

$$S = \left\{\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 : y = mx + b\right\}$$

for some fixed real numbers m and b. Show that the set

$$F(S) := \{F(\vec{v}) : \vec{v} \in S\}$$

is either a line or a point.

Solution. Let $\vec{v} \in S$, then $\vec{v} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ m \end{pmatrix} x + \begin{pmatrix} 0 \\ b \end{pmatrix}$ for some $x \in \mathbb{R}$. Now

$$\begin{aligned} F(\vec{v}) &= F\left(\begin{pmatrix} 1 \\ m \end{pmatrix} x + \begin{pmatrix} 0 \\ b \end{pmatrix}\right) \\ &= F\left(\begin{pmatrix} 1 \\ m \end{pmatrix} x\right) + F\left(\begin{pmatrix} 0 \\ b \end{pmatrix}\right) && \text{using the linearity of } F \\ &= F\left(\begin{pmatrix} 1 \\ m \end{pmatrix}\right) x + F\left(\begin{pmatrix} 0 \\ b \end{pmatrix}\right) && \text{using the linearity of } F. \end{aligned}$$

But this is again the equation of a line, or a point (if $F\left(\begin{pmatrix} 1 \\ m \end{pmatrix}\right) = \vec{0}$).

Activity 5.3. Suppose that $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation given by

$$F(\vec{e}_1) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ and } F(\vec{e}_2) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

(1) Find $F\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$ and $F\left(\begin{pmatrix} 2 \\ 3 \end{pmatrix}\right)$.

(2) Find a formula for $F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right)$.

(3) Find a 2×2 matrix A so that $F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = A\begin{pmatrix} x \\ y \end{pmatrix}$.

Solution.

(1) We know that $\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1\vec{e}_1 + 1\vec{e}_2$, and $\begin{pmatrix} 2 \\ 3 \end{pmatrix} = 2\vec{e}_1 + 3\vec{e}_2$. Since F is a linear transformation,

$$\begin{aligned} F\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) &= F\left(1\vec{e}_1 + 1\vec{e}_2\right) \\ &= F(\vec{e}_1) + F(\vec{e}_2) \\ &= \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ 1 \end{pmatrix}. \end{aligned}$$

And

$$\begin{aligned} F\left(\begin{pmatrix}2\\3\end{pmatrix}\right) &= F\left(2\vec{e}_1+3\vec{e}_2\right)\\ &= 2F(\vec{e}_1)+3F(\vec{e}_2)\\ &= 2\begin{pmatrix}1\\-1\end{pmatrix}+3\begin{pmatrix}1\\2\end{pmatrix}\\ &= \begin{pmatrix}2\\-2\end{pmatrix}+\begin{pmatrix}3\\6\end{pmatrix}\\ &= \begin{pmatrix}5\\4\end{pmatrix}. \end{aligned}$$

(2) Since $\begin{pmatrix}x\\y\end{pmatrix} = x\vec{e}_1 + y\vec{e}_2$, and F is a linear transformation, we have

$$\begin{aligned} F\left(\begin{pmatrix}x\\y\end{pmatrix}\right) &= F(x\vec{e}_1+y\vec{e}_2)\\ &= xF(\vec{e}_1)+yF(\vec{e}_2)\\ &= x\begin{pmatrix}1\\-1\end{pmatrix}+y\begin{pmatrix}1\\2\end{pmatrix}\\ &= \begin{pmatrix}x+y\\-x+2y\end{pmatrix}. \end{aligned}$$

(3) We find the matrix $A = \begin{pmatrix}1 & 1\\-1 & 2\end{pmatrix}$. Note that

$$A\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}1 & 1\\-1 & 2\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}x+y\\-x+2y\end{pmatrix} = F\left(\begin{pmatrix}x\\y\end{pmatrix}\right).$$

Activity 5.4. Find the defining matrices for the following linear transformations.

(1) $F : \mathbb{R}^3 \to \mathbb{R}^3$, $\vec{x} \mapsto 2\vec{x}$

(2) $G : \mathbb{R}^2 \to \mathbb{R}^2$ which rotates every vector $90°$ counterclockwise about the origin.

Solution.

(1) From Theorem 5.4, we have $A_F = \begin{pmatrix}F(\vec{e}_1) & F(\vec{e}_2)\end{pmatrix}$. Now $F(\vec{e}_1) = 2\vec{e}_1 = \begin{pmatrix}2\\0\end{pmatrix}$, and $F(\vec{e}_2) = 2\vec{e}_2 = \begin{pmatrix}0\\2\end{pmatrix}$, and so

$$A_F = \begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}.$$

(2) From Theorem 5.4, we have $A_G = \begin{pmatrix}G(\vec{e}_1) & G(\vec{e}_2)\end{pmatrix}$. Rotating $\vec{e}_1$, $90°$ counterclockwise about the origin results in the vector $G(\vec{e}_1) = \vec{e}_2$. And rotating $\vec{e}_2$, $90°$ counterclockwise about the origin results in the vector $G(\vec{e}_2) = -\vec{e}_1$. So we have

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

Activity 5.5. Recall from Activity 5.1 that the function $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$T\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x+1 \\ y-1 \end{pmatrix}$$

is not linear. Determine what's wrong with the following argument.
We have

$$T(\vec{e}_1) = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, T(\vec{e}_2) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \text{and } T(\vec{e}_3) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

So, $T(\vec{x}) = A\vec{x}$ where

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

By Proposition 5.3, every matrix transformation is linear, and so T is a linear transformation.

Solution. The mistake is that T is not a matrix transformation, and $T(\vec{x}) \neq A\vec{x}$. Theorem 5.4 can only be applied to *linear* transformations, and T is *not* linear.

Activity 5.6. Show that the function $H : \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$H\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+y \\ x \\ 0 \end{pmatrix}$$

is linear by showing that it's equal to a matrix transformation.

Solution. Let $A_H = \begin{pmatrix} H(\vec{e}_1) & H(\vec{e}_2) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}$. Let $\vec{v} = \begin{pmatrix} x \\ y \end{pmatrix}$, let's check if $H(\vec{v}) = A_H\vec{v}$. We have

$$A_H\vec{v} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ x \\ 0 \end{pmatrix} = H\left(\begin{pmatrix} x \\ y \end{pmatrix}\right).$$

Thus, H is a matrix transformation and by Proposition 5.3, it is a linear transformation.

Activity 5.7. Consider the function

$$F : \mathbb{R}^3 \to \mathbb{R}^2, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x-y \\ z \end{pmatrix}.$$

Find a vector $\vec{x}$ in $\ker(F)$. Then, find a vector $\vec{y}$ in $\operatorname{im}(F)$.

Solution. Let $\vec{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \ker(F)$, then $F(\vec{x}) = \vec{0}$. So we must have $F\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x-y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. In other words, for $\vec{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \ker(F)$, we must have $x - y = 0$ or $x = y$, and $z = 0$. The vector $\vec{x} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ is one of the vectors in $\ker(F)$.

Next, to find a vector $\vec{y} \in \operatorname{im}(F)$, we can pick any random vector from the domain of F, and its image will be a vector in $\operatorname{im}(F)$. For example, we can take

$$\vec{y} = F\left(\begin{pmatrix}1\\2\\3\end{pmatrix}\right) = \begin{pmatrix}-1\\3\end{pmatrix},$$

then $\vec{y} \in \operatorname{im}(F)$.

Activity 5.8. Show that for a linear transformation F, we have $\ker(F) = \operatorname{Nul}(A_F)$ and $\operatorname{im}(F) = \operatorname{Col}(A_F)$.

Solution. Let F be a linear transformation with defining matrix A_F. Note that

$$\begin{aligned}\ker(F) &= \{\vec{x} \in \mathbb{R}^n \mid F(\vec{x}) = \vec{0}\}, \text{ by definition of the kernel}\\ &= \{\vec{x} \in \mathbb{R}^n \mid A_F\vec{x} = \vec{0}\}, \text{ since } A_F \text{ is the defining matrix of } F\\ &= \operatorname{Nul}(A_F), \text{ by definition of the null space.}\end{aligned}$$

Next, to see that $\operatorname{im}(F) = \operatorname{Col}(A_F)$. Let $A_F = \begin{pmatrix}\vec{v}_1 & \dots & \vec{v}_n\end{pmatrix}$ note that

$$\begin{aligned}\operatorname{im}(F) &= \{\vec{y} \in \mathbb{R}^m \mid F(\vec{x}) = \vec{y} \text{ for some } \vec{x} \in \mathbb{R}^n\}, \text{ by definition of the image}\\ &= \{\vec{y} \in \mathbb{R}^m \mid A_F\vec{x} = \vec{y} \text{ for some } \vec{x} \in \mathbb{R}^n\}, \text{ since } A_F \text{ is the defining matrix of } F\\ &= \left\{\vec{y} \in \mathbb{R}^m \mid \vec{y} = \begin{pmatrix}\vec{v}_1 & \dots & \vec{v}_n\end{pmatrix}\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix} \text{ for some } x_1, \dots, x_n \in R\right\}\\ &= \{\vec{y} \in \mathbb{R}^m \mid \vec{y} = x_1\vec{v}_1 + \dots + x_n\vec{v}_n \text{ for some } x_1, \dots, x_n \in R\}\\ &= \{\vec{y} \in \mathbb{R}^m \mid \vec{y} \in \operatorname{Span}(\vec{v}_1, \dots, \vec{v}_n)\}\\ &= \operatorname{Col}(A), \text{ by definition of the column space.}\end{aligned}$$

Activity 5.9. Find the rank and nullity of the linear transformation from Activity 5.7. *Solution.* Observe that this function has defining matrix

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

To find $\operatorname{rank}(A)$, let's find a basis for $\operatorname{im}(T_A)$. Observe that we have

$$\begin{aligned}\operatorname{im}(T_A) &= \{\vec{y} \in \mathbb{R}^2 \mid A\vec{x} = \vec{y} \text{ for some } \vec{x} \in \mathbb{R}^3\}\\ &= \{A\vec{x} \mid \vec{x} \in \mathbb{R}^3\}\\ &= \left\{x_1\begin{pmatrix}1\\0\end{pmatrix} + x_2\begin{pmatrix}-1\\0\end{pmatrix} + x_3\begin{pmatrix}0\\1\end{pmatrix} \mid x_1, x_2, x_3 \in \mathbb{R}\right\}\\ &= \operatorname{Span}\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}-1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right).\end{aligned}$$

Since the pivot columns of A are in the first and third column, our work above tells us that we can take

$$\left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}$$

as a basis for $\operatorname{im}(T_A)$. Hence, $\operatorname{rank}(A) = \dim(\operatorname{im}(T_A)) = 2$.

Next, observe that

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \ker(F) \text{ if and only if } A\vec{x} = \vec{0}.$$

This vector equation has corresponding augmented matrix

$$\begin{pmatrix} 31 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

and so x_2 is free and so

$$\begin{aligned} \ker(F) &= \left\{ \begin{pmatrix} x_2 \\ x_2 \\ 0 \end{pmatrix} : x_2 \in \mathbb{R} \right\} \\ &= \left\{ x_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} : x_2 \in \mathbb{R} \right\} \\ &= \operatorname{Span}\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right). \end{aligned}$$

Since this generating set only has one element, it must be a linearly independent set, and hence $\text{nullity}(F) = \dim(\ker(F)) = 1$.

B.6. Chapter 6 Activity Solutions

Activity 6.1. Determine whether each of the following functions are injective, surjective, or bijective.

(1) $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ where
$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

(2) $T_B : \mathbb{R}^2 \to \mathbb{R}^2$ where
$$B = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

(3) $T_C : \mathbb{R}^3 \to \mathbb{R}^2$ where
$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

(4) $T_D : \mathbb{R}^2 \to \mathbb{R}^3$ where
$$D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Solution.

(1) We can express T_A as follows
$$T_A\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$
So T_A is the identity function. This function is bijective. Note that if $T_A\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) = T_A\left(\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right)$, then we must have $T_A\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = T_A\left(\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right)$, and so T_A is injective. Also, for any vector $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$ (which is the codomain of T_A), we can find a vector in the domain of T_A, namely the vector $\begin{pmatrix} x \\ y \end{pmatrix}$, such that $T_A\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x \\ y \end{pmatrix}$. And so T_A is surjective.

(2) We can express T_B as follows
$$T_B\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ 0 \end{pmatrix}.$$
Note that T_B is not injective, because $T_B\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = T_B\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$. It is also not surjective because the second entry in any vector in the image of T_B must be 0. For example, $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is in the codomain of T_B, but it is not in the image of T_B.

(3) We can express T_C as follows
$$T_A\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y+z \end{pmatrix}.$$

Note that T_C is not injective, because $T_C\left(\begin{pmatrix}0\\1\\0\end{pmatrix}\right) = \begin{pmatrix}0\\1\end{pmatrix} = T_C\left(\begin{pmatrix}0\\0\\1\end{pmatrix}\right)$.
However, T_C is surjective. Let $\begin{pmatrix}u_1\\u_2\end{pmatrix} \in \mathbb{R}^2$, then $T_C\left(\begin{pmatrix}u_1\\u_2\\0\end{pmatrix}\right) = \begin{pmatrix}u_1\\u_2\end{pmatrix}$, showing that T_C is onto.

(4) We can express T_D as follows

$$T_D\left(\begin{pmatrix}x\\y\end{pmatrix}\right) = \begin{pmatrix}1 & 0\\0 & 1\\0 & 0\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}x\\y\\0\end{pmatrix}.$$

Note that T_D is not surjective because the third entry in any vector in the image of T_D must be 0. For example, $\begin{pmatrix}1\\1\\1\end{pmatrix}$ is in the codomain of T_D, but it is not in the image of T_D. However, T_D is injective. Suppose $T_D\left(\begin{pmatrix}x_1\\y_1\end{pmatrix}\right) = T_D\left(\begin{pmatrix}x_2\\y_2\end{pmatrix}\right)$. Then $T_D\left(\begin{pmatrix}x_1\\y_1\end{pmatrix}\right) = \begin{pmatrix}x_1\\y_1\\0\end{pmatrix} = \begin{pmatrix}x_2\\y_2\\0\end{pmatrix} = T_D\left(\begin{pmatrix}x_2\\y_2\end{pmatrix}\right)$. In particular, $x_1 = x_2$, and $y_1 = y_2$, which shows that T_D is one-to-one.

Activity 6.2. Determine whether each of the following functions are injective, surjective, or bijective.

(1) $F : \mathbb{R}^2 \to \mathbb{R}^3$, $\begin{pmatrix}x\\y\end{pmatrix} \mapsto \begin{pmatrix}x\\y\\0\end{pmatrix}$

(2) $G : \mathbb{R}^3 \to \mathbb{R}^3$, $\begin{pmatrix}x\\y\\z\end{pmatrix} \mapsto \begin{pmatrix}x-y\\y+z\\x+z\end{pmatrix}$

(3) $H : \mathbb{R}^3 \to \mathbb{R}^3$, $\begin{pmatrix}x\\y\\z\end{pmatrix} \mapsto \begin{pmatrix}x-y\\y+z\\z\end{pmatrix}$

Solution.

(1) We claim that F is injective but not surjective. To see this, observe that F has defining matrix

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

By Theorem 6.3, F is injective because every column in $\text{rref}(A_F)$ has a pivot. Also, F is not surjective because the third row in $\text{rref}(A)$ does not have a pivot.

(2) We claim that G is neither surjective nor injective. To see this, observe that G has defining matrix
$$B = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$
By Theorem 6.3, G is not injective because there is no pivot in the third column of $\operatorname{rref}(B)$. Also, G is not surjective because there is no pivot in the third row of $\operatorname{rref}(B)$.

(3) We claim that H is bijective. To see this, observe that H has defining matrix
$$C = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$
By Theorem 6.3, H is injective because $\operatorname{rref}(C)$ has a pivot in every column. It is also surjective, because $\operatorname{rref}(C)$ has a pivot in every row.

Activity ??. Find I_2, I_3 and I_4. Then, give a description for what I_n looks like for a general integer n. *Solution.* The defining matrix for $\mathrm{id}_{\mathbb{R}^2}(\vec{x}) = \vec{x}$ is the matrix $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The defining matrix for $\mathrm{id}_{\mathbb{R}^3}(\vec{x}) = \vec{x}$ is the matrix $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.
The defining matrix for $\mathrm{id}_{\mathbb{R}^4}(\vec{x}) = \vec{x}$ is the matrix $I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.
In general, I_n is the $n \times n$ matrix with zeros everywhere, except the diagonal entries, which are 1.

Activity 6.4. Consider the matrix
$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$
Verify that
$$A^{-1} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$
Solution. We just need to check that $AA^{-1} = A^{-1}A = I_3$.
$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Activity ??. Use Theorem 6.17 to find the inverse of
$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & -1 \end{pmatrix}.$$

Solution. We start with $\left(A \mid I_n\right)$ and row reduce until we get $\left(I_n \mid A^{-1}\right)$, as follows:

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 4 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & -5 & 1 & -2 & 0 \\ 0 & 1 & 4 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{array}\right) \text{ by subtracting } 2R_2 \text{ from } R_1$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & -5 & 1 & -2 & 0 \\ 0 & 1 & 4 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{array}\right) \text{ by multiplying } R_3 \text{ by } -1$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -2 & -5 \\ 0 & 1 & 4 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{array}\right) \text{ by adding } 5R_3 \text{ to } R_1$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -2 & -5 \\ 0 & 1 & 0 & 0 & 1 & 4 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{array}\right) \text{ by subtracting } 4R_3 \text{ from } R_2$$

And this gives us $A^{-1} = \begin{pmatrix} 1 & -2 & -5 \\ 0 & 1 & 4 \\ 0 & 0 & -1 \end{pmatrix}$.

Activity ??. Consider the matrix

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Calculate the following matrix products $S_{1,3}B$, $M_2(5)B$, and $A_{1,2}(5)B$. What do you notice? *Solution.* We get the following,

$$S_{1,3}B = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 7 & 8 & 9 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{pmatrix},$$

$$M_2(5)B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 20 & 25 & 30 \\ 7 & 8 & 9 \end{pmatrix},$$

$$A_{1,2}(5)B = \begin{pmatrix} 1 & 5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 21 & 27 & 33 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Notice that $S_{1,3}B$ is just B with the first and third rows swapped. While $M_2(5)B$ is B with its second column multiplied by 5. Finally, $A_{1,2}(5)B$ is B with twice the second row added to the first row. So when the elementary matrices are multiplied on the left of B, the result is the matrix B after performing the same elementary row operation that gave rise to the elementary matrix!

Activity 6.7. Let's show that every elementary matrix is invertible.

(1) Find the inverse of the 3×3 elementary matrices

$$S_{1,3}, M_2(5), \text{ and } A_{1,2}(5).$$

(2) Find the inverse of the $n \times n$ elementary matrices

$$S_{ij}, M_i(c), \text{ and } A_{i,j}(c),$$

where c is any nonzero real number.

Solution.

(1) We've seen that the elementary matrices 'perform' a row operation on a matrix when multiplied by it. We want to find the matrix $S_{1,3}^{-1}$, so that $S_{1,3}S_{1,3}^{-1} = I_3$. So swapping the first and third rows of $S_{1,3}^{-1}$ should give us the identity matrix. But this just means that $S_{1,3}^{-1}$ should be the identity matrix with its first and third rows swapped. In other words, the inverse of $S_{1,3}$ is $S_{1,3}$ itself. (We could also compute the inverse directly using Theorem 6.17.)

We want to find the matrix $M_2(5)^{-1}$, so that $M_2(5)M_2(5)^{-1} = I_3$. So multiplying the second row of $M_2(5)^{-1}$ by 5 should give us the identity matrix. But this just means that $M_2(5)^{-1}$ should be the identity matrix with its second row divided by 5. In other words, the inverse of $M_2(5)$ is $M_2(\frac{1}{5})$. (We could also compute the inverse directly using Theorem 6.17.)

Let's use Theorem 6.17 to find the inverse of $A_{1,2}(5)$,

$$\left(\begin{array}{ccc|ccc} 1 & 5 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -5 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \text{by subtracting } 5R_2 \text{ from } R_1$$

We can see that the inverse of $A_{1,2}(5)$ is simply the elementary matrix $A_{1,2}(-5)$.

(2) From our arguments above, we can see that the inverse of $S_{i,j}$ is $S_{i,j}$ itself, because swapping two rows is 'undone' by swapping the same two rows again. The inverse of $M_i(c)$ is $M_i(\frac{1}{c})$, because multiplying a row by a constant is 'undone' by dividing the same row by that constant. And the inverse of $A_{i,j}(c)$ is $A_{i,j}(-c)$, because adding a multiple of a row j to row i is 'undone' by subtracting the same multiple row j from row i.

B.7. Chapter 7 Activity Solutions

Activity 7.1. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by

$$F(\vec{e}_1) = 3\vec{e}_1 \text{ and } F(\vec{e}_2) = 2\vec{e}_2.$$

(1) Sketch a picture of the unit square S.

(2) Sketch a picture of $F(S) := \{F(\vec{v}) : \vec{v} \in S\}$ as a subset of $\mathbb{R}^2$.

(3) Sketch the image of the "standard coordinate grid" for $\mathbb{R}^2$ under F.

Solution.

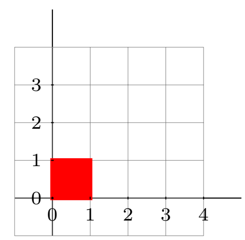


Figure 1. The unit square S.

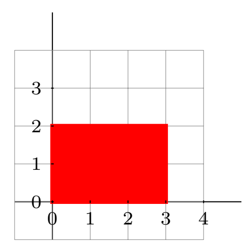


Figure 2. The image of the unit square under F.

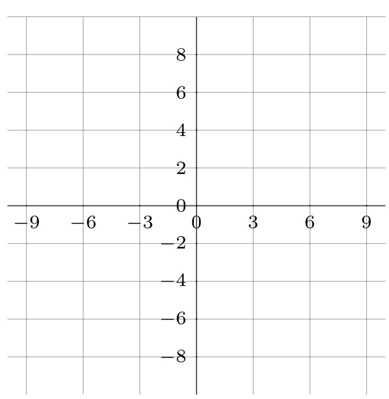


Figure 3. The image of the standard coordinate grid under F.

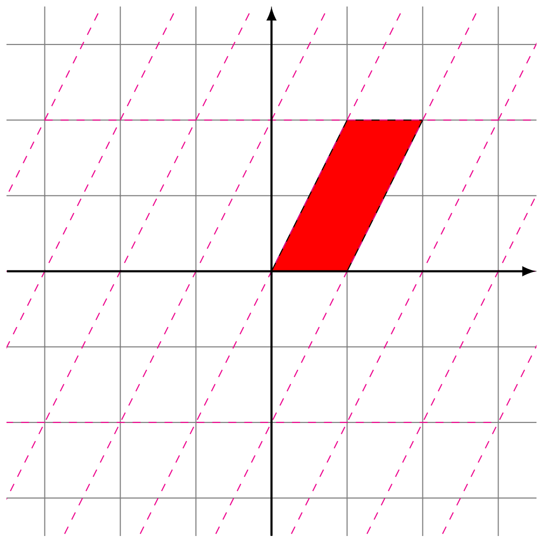
Activity 7.2. Let $G : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by

$$G(\vec{e}_1) = \vec{e}_1 + \vec{e}_2 \text{ and } G(\vec{e}_2) = 2\vec{e}_2.$$

(1) Sketch a picture of $G(S)$ as a subset of $\mathbb{R}^2$.

(2) Sketch the image of the "standard coordinate grid" for $\mathbb{R}^2$ under G.

Solution. The standard coordinate grid is drawn in solid black lines below, and its image is drawn in red dashed lines.



Activity 7.3. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation.

(1) Suppose that F is surjective. Use geometric reasoning to argue that the area of $F(S)$ is positive.

(2) Suppose that F is not surjective. Use geometric reasoning to argue that the area of $F(S)$ is equal to zero.

(3) Suppose that the area of $F(S)$ is equal to 6. What will the image of the standard coordinate grid under F look like?

(4) Suppose that the area of $F(S)$ is equal to 1/10. What will the image of the standard coordinate grid under F look like?

Solution.

(1) We claim that the images of $\vec{e}_1$ and $\vec{e}_2$ are linearly independent. If they are not, then $F(\vec{e}_2) = cF(\vec{e}_1)$ for some scalar c. But then $\operatorname{im}(F) = \operatorname{Span}(F(\vec{e_1}))$, which cannot be all of $\mathbb{R}^2$ contradicting the surjectivity of F. Since $F(\vec{e}_1)$ and $F(\vec{e}_2)$ are linearly independent, they are not on the same line, and so they are two sides of a parallelogram, which has a positive area. By Proposition 7.2, $F(S)$ is the parallelogram with sides $F(\vec{e}_1)$ and $F(\vec{e}_2)$, which must have a positive area.

(2) We claim that $F(\vec{e}_2) = cF(\vec{e}_1)$ for some scalar c. Otherwise, $F(\vec{e}_1)$ and $F(\vec{e}_2)$ are linearly independent, and $\operatorname{im}(F) = \operatorname{Span}(\{F(\vec{e}_1), F(\vec{e}_2)\}) = \mathbb{R}^2$ which contradicts that F is not surjective. Now since $F(\vec{e}_2) = cF(\vec{e}_1)$, it follows that

$F(\vec{e}_1), F(\vec{e}_2)$ are on the same line. By Proposition 7.2, $F(S)$ is the parallelogram with sides $F(\vec{e}_1)$ and $F(\vec{e}_2)$, which is a line and so has area equal to zero.

(3) It would be a grid with axes in the $F(\vec{e}_1)$ and $F(\vec{e}_2)$ directions. And one parallelogram in this grid would have area equal to 6.

(4) It would be a grid with axes in the $F(\vec{e}_1)$ and $F(\vec{e}_2)$ directions. And one parallelogram in this grid would have area equal to 1/10.

Activity 7.4. Find the orientation for the following ordered bases for $\mathbb{R}^2$.

(1) $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}$ where

$$\vec{b}_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \text{ and } \vec{b}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

(2) $\mathcal{C} = \{\vec{c}_1, \vec{c}_2\}$ where

$$\vec{c}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \vec{c}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Solution.

(1) We can rotate $\vec{b}_1$ less than $180°$ counterclockwise to reach $\vec{b}_2$, and so the basis $\mathcal{B}$ is positively oriented.

(2) We must rotate $\vec{c}_1$ $270° > 180°$ counterclockwise to reach $\vec{c}_2$, and so the basis $\mathcal{C}$ is negatively oriented.

Activity 7.5. Find the determinant of the matrices

$$A = \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Solution. First, let's find $T_A(\vec{e}_1)$ and $T_A(\vec{e}_2)$. We have

$$T_A(\vec{e}_1) = \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, T_A(\vec{e}_2) = \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Note that we can rotate $T_A(\vec{e}_1)$ less than $180°$ counterclockwise to reach $\vec{e}_2$, so $\{T_A(\vec{e}_1), T_A(\vec{e}_2)\}$ is positively oriented. Now the determinant of A is the area of the parallelogram with sides $T_A(\vec{e}_1)$ and $T_A(\vec{e}_2)$. Recall that the area of a parallelogram is equal to the langth of its base multiplied by the length of its height. This parallelogram has a base of length 2, and its height is 1, and so its area is 2×1. Therefore, $\det(A) = 2$.

Next, we find $\det(B)$. Let's find $T_B(\vec{e}_1)$ and $T_B(\vec{e}_2)$. We have

$$T_B(\vec{e}_1) = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, T_B(\vec{e}_2) = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Note that we must rotate $T_B(\vec{e}_1)$ $270° > 180°$ counterclockwise to reach $T_B(\vec{e}_2)$, so $\{T_A(\vec{e}_1), T_A(\vec{e}_2)\}$ is negatively oriented. Now the determinant of B is the area of the parallelogram with sides $T_B(\vec{e}_1)$ and $T_B(\vec{e}_2)$ multiplied by -1. Since this parallelogram is just the unit square rotated $45°$ clockwise, its area remains 1. Therefore, $\det(B) = -1$.

Activity 7.6. In this activity, we develop a method to calculate the determinant of a 2×2 matrix completely algebraically. Let

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

For simplicity we'll assume that the set $\left\{ \begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \right\}$ is a positively oriented ordered basis with $d \neq 0$ and both vectors in the first quadrant.

(1) Recall that the area of a parallelogram can be computed as the product of its base times its height. Use this observation to calculate the determinant of

$$\begin{pmatrix} e & b \\ 0 & d \end{pmatrix}.$$

(2) Show that

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det \begin{pmatrix} a - \frac{bc}{d} & b \\ 0 & d \end{pmatrix}.$$

(3) Use the previous two parts to conclude that

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Solution.

(1) First, we find $T_A(\vec{e}_1)$, and $T_A(\vec{e}_2)$. We have

$$T_A(\vec{e}_1) = \begin{pmatrix} e & b \\ 0 & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} e \\ 0 \end{pmatrix}, T_A(\vec{e}_2) = \begin{pmatrix} e & b \\ 0 & d \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}.$$

Since $F(\vec{e}_2) = \begin{pmatrix} b \\ d \end{pmatrix}$ is in the first quadrant, and $F(\vec{e}_1) = \begin{pmatrix} e \\ 0 \end{pmatrix}$ is on the $x - axis$, it follows that $\{F(\vec{e}_1), F(\vec{e}_2)\}$ is a positively oriented basis (as long as $e \neq 0$). The two vectors $F(\vec{e}_1), F(\vec{e}_2)$ are the sides of a parallelogram with base length equal to e, and height equal to d. Therefore, $\det(A) = ed$.

(2)

(3) We have the following,

$$\begin{aligned} \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \det \begin{pmatrix} a - \frac{bc}{d} & b \\ 0 & d \end{pmatrix} && \text{by part (2)} \\ &= (a - \frac{bc}{d})d && \text{by part(1)} \\ &= ad - bc. \end{aligned}$$

Activity 7.7. Use geometric reasoning to argue that a 2×2 matrix A is invertible if and only if $\det(A) \neq 0$. Conclude that $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. (Note that this gives an alternate method to prove Chapter Exercise P6.3).

Solution. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Suppose A is invertible, then T_A is surjective by an exercise in chapter 6. By activity 7.3, we know that $T_A(S)$ will have a positive area.

In other words, we will have $\det(A) \neq 0$. For the converse, assume that $\det(A) \neq 0$. Then $ad - bc \neq 0$. Now suppose the columns of A are linearly dependent, then we must have $\begin{pmatrix} b \\ d \end{pmatrix} = k \begin{pmatrix} a \\ c \end{pmatrix}$ for some scalar k. So that $b = ka$, and $d = kc$. But then $ad - bc = a(kc) - (ka)c = 0$ contradicting our assumption that $\det(A) \neq 0$. So the columns of A must be linearly independent. By an exercise in chapter 6, it follows that A is invertible.

Activity 7.8. Let $H : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation.

(1) Sketch a picture of the unit cube C as a subset of $\mathbb{R}^3$.

(2) What type of geometric object could the set $H(C_3)$ be equal to?

Solution.

(2) The set $H(C_3)$ could be a volume (a 3-dimensional subset of $\mathbb{R}^3$), or a subset of a plane, or a line segment, or just a point (the zero vector).

Activity 7.9. Find the orientation of the following ordered bases for $\mathbb{R}^3$.

(1) $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \vec{b}_3\}$ where

$$\vec{b}_1 = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}, \vec{b}_2 = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \text{ and } \vec{b}_3 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}.$$

(2) $\mathcal{C} = \{\vec{c}_1, \vec{c}_2, \vec{c}_3\}$ where

$$\vec{c}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \vec{c}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \vec{c}_3 = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}.$$

Solution.

(1) The basis $\mathcal{B}$ is positively oriented because it satisfies the right-hand rule.

(2) The basis $\mathcal{C}$ is negatively oriented because it does not satisfy the right-hand rule.

Activity 7.10. Calculate the determinant of the matrices

$$A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 1 \end{pmatrix}, \text{ and } B = \begin{pmatrix} 0 & 0 & 2 \\ 1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix}.$$

Solution. By Proposition 7.9, we have

$$\begin{aligned} \det(A) &= \det \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 1 \end{pmatrix} \\ &= 3 \det \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix} && \text{by Proposition 7.9} \\ &= 3(2 \cdot 1 - 2 \cdot 0) && \text{by Proposition ??} \\ &= 3 \cdot 2 \\ &= 6. \end{aligned}$$

By Proposition 7.10, we have

$$\begin{aligned}
\det(B) &= \det\begin{pmatrix} 0 & 0 & 2 \\ 1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix} \\
&= -\det\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{pmatrix} && \text{by Proposition 7.10} \\
&= -2\det\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} && \text{by Proposition 7.9} \\
&= -2(1\cdot 1 - 1\cdot(-1)) && \text{by Proposition ??} \\
&= -2\cdot 2 \\
&= -4
\end{aligned}$$

Activity 7.11. Use the method in the previous example to calculate $\det(A)$, where

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

Given this calculation, what can you say about the matrix transformation $T_A : \mathbb{R}^3 \to \mathbb{R}^3$?

Solution. We proceed as in the example above. First, we use Proposition 7.11 to get

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

Next, we use Proposition 7.10, which yields

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

Finally, we can use Proposition 7.9 to get

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

Since the determinant of A is not equal to zero, we can say that the matrix transformation $T_A : \mathbb{R}^3 \to \mathbb{R}^3$ is surjective and therefore, invertible.

Activity 7.12. Find the determinant of the following matrix

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

Solution. Using the definition, as well as Proposition 7.9, and Proposition 7.10, we get

$$\begin{aligned}
\det(A) &= 1\cdot\det\begin{pmatrix}2&0&0\\0&-1&0\\1&1&2\end{pmatrix} - 0\cdot\det\begin{pmatrix}0&0&0\\1&-1&0\\0&1&2\end{pmatrix} + 1\cdot\det\begin{pmatrix}0&2&0\\1&0&0\\0&1&2\end{pmatrix} - 0\cdot\det\begin{pmatrix}0&2&0\\1&0&-1\\0&1&1\end{pmatrix}\\
&= \det\begin{pmatrix}2&0&0\\0&-1&0\\1&1&2\end{pmatrix} + \det\begin{pmatrix}0&2&0\\1&0&0\\0&1&2\end{pmatrix}\\
&= \left(2\det\begin{pmatrix}-1&0\\1&2\end{pmatrix}\right) + \left(-\begin{pmatrix}2&0&0\\0&1&0\\1&0&2\end{pmatrix}\right)\\
&= 2\cdot(-2) - 2\cdot\begin{pmatrix}1&0\\0&2\end{pmatrix}\\
&= -4 - 2\cdot 2\\
&= -8.
\end{aligned}$$

Activity 7.13.

(1) Recall the matrix A from 7.11

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

Use cofactor expansion along the first column of A to verify Lemma 7.19 holds for this example.

(2) Verify Lemma 7.19 holds for an arbitrary 3×3 matrix

$$A = \begin{pmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{pmatrix}.$$

Solution.

(1) Using cofactor expansion along the first column of A, we get

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

This is the same answer we got for 7.11.

(2) For an arbitrary 3×3 matrix, Lemma 7.19 (cofactor expansion along the first column) gives

$$\begin{aligned}
\det(A) &= a_{11}\cdot\det\begin{pmatrix}a_{22}&a_{23}\\a_{32}&a_{33}\end{pmatrix} - a_{21}\cdot\det\begin{pmatrix}a_{12}&a_{13}\\a_{32}&a_{33}\end{pmatrix} + a_{31}\cdot\det\begin{pmatrix}a_{12}&a_{13}\\a_{22}&a_{23}\end{pmatrix}\\
&= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{21}(a_{12}a_{33} - a_{13}a_{32}) + a_{31}(a_{12}a_{23} - a_{13}a_{22})\\
&= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31}.
\end{aligned}$$

And using cofactor expansion along the first row, we get

$$\begin{aligned}\det(A) &= a_{11}\cdot\det\begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33}\end{pmatrix} - a_{12}\cdot\det\begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33}\end{pmatrix} + a_{13}\cdot\det\begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32}\end{pmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.\end{aligned}$$

But this is the same expression we got from applying Lemma 7.19.

Activity ??. Use Lemma 7.19 to show the following.

(1) Verify Theorem 7.20 holds for an arbitrary 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{pmatrix}.$$

(2) Verify Theorem 7.20 holds for an arbitrary 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{pmatrix}.$$

Solution.

(1) First note that $\det(A) = a_{11}a_{22} - a_{12}a_{21}$. We need to check three things. First we check that interchanging two rows of A multiplies the determinant by -1. Let $B = \begin{pmatrix} a_{21} & a_{22} \\ a_{11} & a_{12}\end{pmatrix}$, the matrix obtained by interchanging the two rows of A. Then $\det(B) = a_{21}a_{12} - a_{22}a_{11} = -a_{11}a_{22} + a_{12}a_{21} = -\det(A)$.

Now let's see what happens when we multiply a row of A by a scalar c. Let $B = \begin{pmatrix} ca_{11} & ca_{12} \\ a_{21} & a_{22}\end{pmatrix}$. Then $\det(B) = ca_{11}a_{22} - ca_{12}a_{21} = c(a_{11}a_{22} - a_{12}a_{21}) = c\det(A)$. Similarly, if $D = \begin{pmatrix} a_{11} & a_{12} \\ ca_{21} & ca_{22}\end{pmatrix}$, then $\det(D) = c\det(A)$.

Lastly, we need to show that adding a scalar multiple of one row to another does not affect the determinant. Let k be a scalar, and let B be the matrix obtained from A by replacing R_1 by $R_1 + kR_2$. Then $B = \begin{pmatrix} a_{11} + ka_{21} & a_{12} + ka_{22} \\ a_{21} & a_{22}\end{pmatrix}$, and

$$\begin{aligned}\det(B) &= (a_{11} + ka_{21})a_{22} - (a_{12} + ka_{22})a_{21} \\ &= a_{11}a_{22} + ka_{21}a_{22} - a_{12}a_{21} - ka_{22}a_{21} \\ &= a_{11}a_{22} - a_{12}a_{21} \\ &= \det(A).\end{aligned}$$

Similarly, if D is the matrix obtained from A by replacing R_2 by $R_2 + kR_1$, we get $\det(D) = \det(A)$.

This shows that Theorem 7.20 is True for 2×2 matrices.

(2) We'll show one example of the effect of each class of row operations on the determinant of a matrix, and the rest can be verified in a similar way.

First, let $B = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, the matrix we get by interchanging the first two rows of A. Then using Lemma 7.19, we get

$$\begin{aligned} \det(B) &= a_{21} \cdot \det \begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix} - a_{11} \cdot \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} + a_{31} \cdot \det \begin{pmatrix} a_{22} & a_{23} \\ a_{12} & a_{13} \end{pmatrix} \\ &= a_{21}(a_{12}a_{33} - a_{13}a_{32}) - a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{31}(a_{13}a_{22} - a_{12}a_{23}) \\ &= a_{12}a_{21}a_{33} - a_{13}a_{21}a_{32} - a_{11}a_{22}a_{33} + a_{11}a_{23}a_{32} + a_{13}a_{22}a_{31} - a_{12}a_{23}a_{31} \\ &= -(a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31}) \\ &= -\det(A). \end{aligned}$$

Where we compared the expression we got for $\det(B)$ to the $\det(A)$ we computed in 7.13. Similarly, we can show that interchanging the first and third rows of A or the last two rows of A will give us the same result.

Next, consider the matrix B obtained from A by multiplying the first row of A by a scalar c, so $B = \begin{pmatrix} ca_{11} & ca_{12} & ca_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. Using Lemma 7.19, we get

$$\begin{aligned} \det(B) &= ca_{11} \cdot \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{21} \cdot \det \begin{pmatrix} ca_{12} & ca_{13} \\ a_{32} & a_{33} \end{pmatrix} + a_{31} \cdot \det \begin{pmatrix} ca_{12} & ca_{13} \\ a_{22} & a_{23} \end{pmatrix} \\ &= ca_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{21}(ca_{12}a_{33} - ca_{13}a_{32}) + a_{31}(ca_{12}a_{23} - ca_{13}a_{22}) \\ &= ca_{11}a_{22}a_{33} - ca_{11}a_{23}a_{32} - ca_{12}a_{21}a_{33} + ca_{13}a_{21}a_{32} + ca_{12}a_{23}a_{31} - ca_{13}a_{22}a_{31} \\ &= c(a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31}) \\ &= c\det(A). \end{aligned}$$

Where we compared the expression we got for $\det(B)$ to the $\det(A)$ we computed in 7.13. Similarly, we can show that multiplying the second or third row of A by a scalar will give us the same result.

Lastly, consider the matrix B obtained from A by replacing R_1 by $R_1 + kR_2$ for some scalar k. Then $B = \begin{pmatrix} a_{11} - ka_{21} & a_{12} - ka_{22} & a_{13} - ka_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ and using Lemma 7.19, we get

$$\begin{aligned} \det(B) &= (a_{11} - ka_{21}) \cdot \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{21} \cdot \det \begin{pmatrix} a_{12} - ka_{22} & a_{13} - ka_{23} \\ a_{32} & a_{33} \end{pmatrix} + a_{31} \cdot \det \begin{pmatrix} a_{12} - ka_{22} & a_{13} - ka_{23} \\ a_{22} & a_{23} \end{pmatrix} \\ &= (a_{11} - ka_{21})(a_{22}a_{33} - a_{23}a_{32}) - a_{21}[(a_{12} - ka_{22})a_{33} - (a_{13} - ka_{23})a_{32}] + a_{31}[(a_{12} - ka_{22})a_{23} - (a_{13} - ka_{23})a_{22}] \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - ka_{21}a_{22}a_{33} + ka_{21}a_{23}a_{32} - a_{21}(a_{12}a_{33} - ka_{22}a_{33} - a_{13}a_{32} + ka_{23}a_{32}) + a_{31}(a_{12}a_{23} - ka_{22}a_{23} - a_{13}a_{22} + ka_{22}a_{23}) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - ka_{21}a_{22}a_{33} + ka_{21}a_{23}a_{32} - a_{12}a_{21}a_{33} + ka_{21}a_{22}a_{33} + a_{13}a_{21}a_{32} - ka_{21}a_{23}a_{32} + a_{12}a_{23}a_{31} - ka_{22}a_{23}a_{31} - a_{13}a_{22} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31} \\ &= \det(A). \end{aligned}$$

Where we compared the expression we got for $\det(B)$ to the $\det(A)$ we computed in 7.13. Similarly, we can show that replacing any row of A by the same row added to a scalar multiple of another row of A will give us the same result.

B.8. Chapter 8 Activity Solutions

Activity 8.1. For each of the following matrix-vector pairs, determine whether $\vec{x}$ is an eigenvector of the given matrix. If it is, find the corresponding eigenvalue λ.

(1) $A = \begin{pmatrix} 3 & 2 \\ 3 & 8 \end{pmatrix}$ and $\vec{x} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$.

(2) $B = \begin{pmatrix} 3 & 2 \\ 3 & 8 \end{pmatrix}$ and $\vec{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

(3) $C = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ and $\vec{x} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

Solution.

(1) Note that $A\vec{x} = \begin{pmatrix} 3 & 2 \\ 3 & 8 \end{pmatrix}\begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} -4 \\ 2 \end{pmatrix} = 2\begin{pmatrix} -2 \\ 1 \end{pmatrix}$. Therefore $\vec{x}$ is an eigenvector of A with eigenvalue $\lambda = 2$.

(2) Note that $B\vec{x} = \begin{pmatrix} 3 & 2 \\ 3 & 8 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 11 \end{pmatrix} \neq \lambda\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ for any real number λ. Therefore, the vector $\vec{x}$ is not an eigenvector for B.

(3) Note that $C\vec{x} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0\begin{pmatrix} -2 \\ 1 \end{pmatrix}$. Therefore, $\vec{x}$ is an eigenvector of C with eigenvalue $\lambda = 0$.

Activity 8.2. Find the 2-eigenspace of $A = \begin{pmatrix} 3 & 2 \\ 3 & 8 \end{pmatrix}$.

Solution. The 2-eigenspace of A is the null space of $A-2I_2$. Now $A-2I_2 = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}$ can be row-reduced to $\begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$. We need to find the solution to the homogeneous system of equations. We find that $x + 2y = 0$, setting y to be a free variable t, we get $x = -2t$. Therefore, the 2-eigenspace of A is $\mathrm{Nul}(A - 2I_2) = \mathrm{Span}\left(\begin{pmatrix} -2 \\ 1 \end{pmatrix}\right)$.

Activity 8.3. Find the characteristic polynomial of the following matrices. Then, use Proposition 8.4 to find the eigenvalues of each matrix.

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}, C = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Solution. For matrix A, we find the characteristic polynomial as follows,

$$\begin{aligned}\chi_A(x) &= \det(A - xI_2)\\ &= \det\begin{pmatrix}1-x & 1\\ 1 & 1-x\end{pmatrix}\\ &= (1-x)(1-x) - 1\cdot 1\\ &= x^2 - 2x + 1 - 1\\ &= x^2 - 2x\\ &= x(x-2).\end{aligned}$$

By Proposition 8.4, the eigenvalues of A are the solutions to the equation $\chi_A(x) = 0$. Now $\chi_A(x) = x(x-2) = 0$, when $x = 0$ or $x = 2$. Therefore, A has eigenvalues $\lambda_1 = 0$, and $\lambda_2 = 2$.

Next we find the characteristic polynomial for B,

$$\begin{aligned}\chi_B(x) &= \det(B - xI_3)\\ &= \det\begin{pmatrix}1-x & 0 & 1\\ 0 & 1-x & 1\\ 0 & 0 & 2-x\end{pmatrix}\\ &= (1-x)(1-x)(2-x).\end{aligned}$$

By Proposition 8.4, the eigenvalues of B are the solutions to the equation $\chi_B(x) = 0$. Now $\chi_B(x) = (1-x)^2(2-x) = 0$, when $x = 1$ or $x = 2$. Therefore, A has eigenvalues $\lambda_1 = 1$, and $\lambda_2 = 2$.

Next we find the characteristic polynomial for C,

$$\begin{aligned}\chi_C(x) &= \det(C - xI_2)\\ &= \det\begin{pmatrix}1-x & -1\\ 1 & 1-x\end{pmatrix}\\ &= (1-x)(1-x) - (-1)\cdot 1\\ &= x^2 - 2x + 1 + 1\\ &= x^2 - 2x + 2\end{aligned}$$

By Proposition 8.4, the eigenvalues of C are the solutions to the equation $\chi_C(x) = 0$. However, $\chi_C(x) = x^2 - 2x + 2 = 0$ has no real solutions. This means that C does not have any real eigenvalues.

Activity 8.4. Find the factorization and roots of the following polynomials. For each root, find its algebraic multiplicity.

(1) $f(x) = x^2 - 2x - 3$

(2) $g(x) = x^4 - 4x^2 + 4$

(3) $h(x) = x^2 + 1$

Solution.

(1) First, we find the factorization,

$$f(x) = (x+1)(x-3).$$

The roots of f are -1 and 3, and each of them has algebraic multiplicity equal to 1.

(2) First, we find the factorization,

$$g(x) = (x^2 - 2)^2 = [(x - \sqrt{2})(x + \sqrt{2})]^2 = (x - \sqrt{2})^2(x + \sqrt{2})^2.$$

The roots of g are $-\sqrt{2}$, and $\sqrt{2}$. Each root has algebraic multiplicity equal to 2.

(3) First, we find the factorization,

$$h(x) = (x + i)(x - i),$$

where $i = \sqrt{-1}$. The roots of h are i and $-i$, and each of the roots has algebraic multiplicity equal to 1.

Activity 8.5. Let

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

Find the eigenvalues of A. For each eigenvalue, find its geometric and algebraic multiplicity.

Solution. First, we find the characteristic polynomial,

$$\begin{aligned} \chi_A(x) &= \det(A - xI_3) \\ &= \det \begin{pmatrix} 1-x & 0 & 1 \\ 0 & 1-x & 1 \\ 0 & 0 & 2-x \end{pmatrix} \\ &= (1-x)(1-x)(2-x) \\ &= (1-x)^2(2-x). \end{aligned}$$

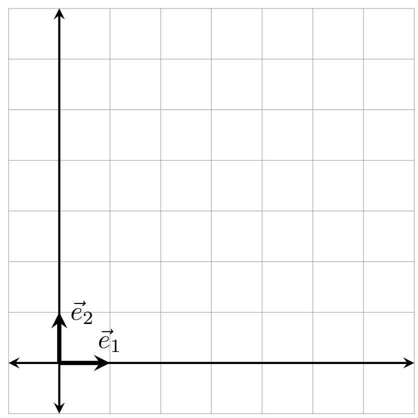
The factorization for $\chi_A(x)$ is shown, and we can see that it has two roots, $\lambda_1 = 1$ has algebraic multiplicity 2 and $\lambda_2 = 2$ has algebraic multiplicity 1. To find the geometric multiplicities, we need to find the λ-eigenspaces.

For $\lambda_1 = 1$, we need to find $\mathrm{Nul}(A - I)$. Now $A - I = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ can be reduced to $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Then $\mathrm{Nul}(A - I) = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$, and so $\lambda_1 = 1$ has geometric multiplicity equal to 2.

For $\lambda_2 = 2$, we need to find $\mathrm{Nul}(A - 2I)$. Now $A - 2I = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ which has reduced row echelon form $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$. Then $\mathrm{Nul}(A - I) = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$, and so $\lambda_2 = 2$ has geometric multiplicity equal to 1.

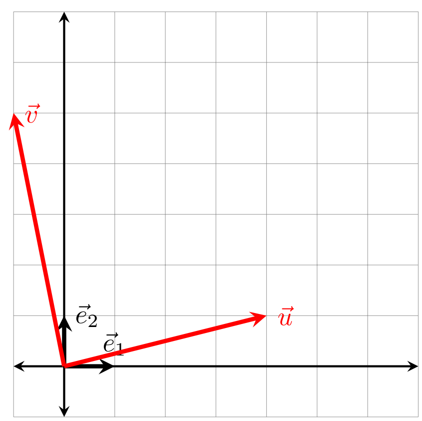
B.9. Chapter 9 Activity Solutions

Activity 9.1. The standard coordinate grid for $\mathbb{R}^2$ is drawn below. Explain how we can use the standard basis $\mathcal{E} = \{\vec{e}_1, \vec{e}_2\}$ to draw the grid lines below.



Then, draw the vectors $\vec{u} = 4\vec{e}_1 + \vec{e}_2$ and $\vec{v} = -\vec{e}_1 + 5\vec{e}_2$ on the coordinate grid above.

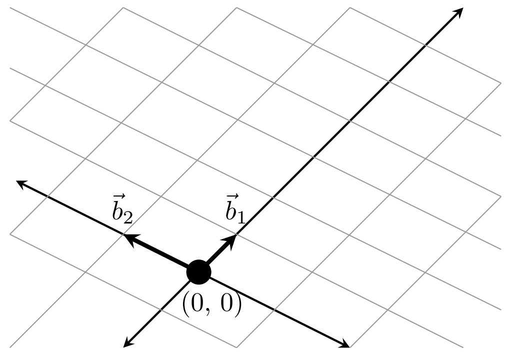
Solution. We draw $\operatorname{Span}(\vec{e}_1)$ and $\operatorname{Span}(\vec{e}_2)$ to get the axes, and we draw $k\vec{e}_1 + c\vec{e}_2$ for integers k and scalars c to get the vertical grid lines. We draw $c\vec{e}_1 + k\vec{e}_2$ for integers k and scalars c to get the horizontal grid lines.



Activity 9.2. Let $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}$

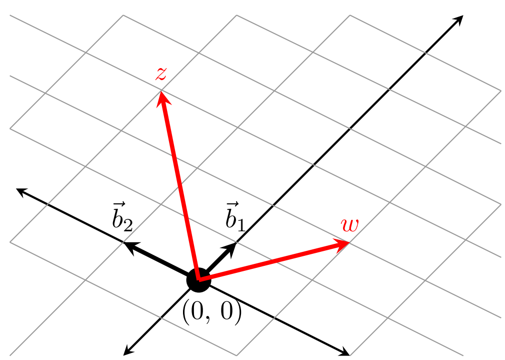
$$\vec{b}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \vec{b}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Observe that $\mathcal{B}$ is a basis for $\mathbb{R}^2$. The image below shows how we can use $\mathcal{B}$ to create a "nonstandard" coordinate grid for $\mathbb{R}^2$. Explain how the vectors $\vec{b}_1, \vec{b}_2$ can be used to draw the grid lines below.



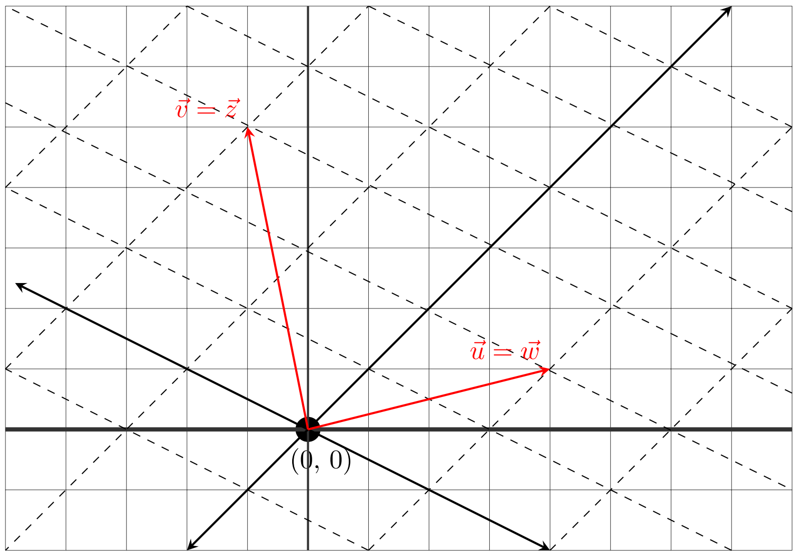
Then, draw the vectors $\vec{w} = 2\vec{b}_1 - \vec{b}_2$ and $\vec{z} = 3\vec{b}_1 + 2\vec{b}_2$ on the coordinate grid above.

Solution. We draw $\text{Span}(\vec{b}_1)$ and $\text{Span}(\vec{b}_2)$ to get the axes, and we draw $k\vec{b_1} + c\vec{b_2}$ for integers k and scalars c to get the grid lines parallel to $\vec{b}_2$. We draw $c\vec{b}_1 + k\vec{b}_2$ for integers k and scalars c to get the grid lines parallel to $\vec{b}_1$.



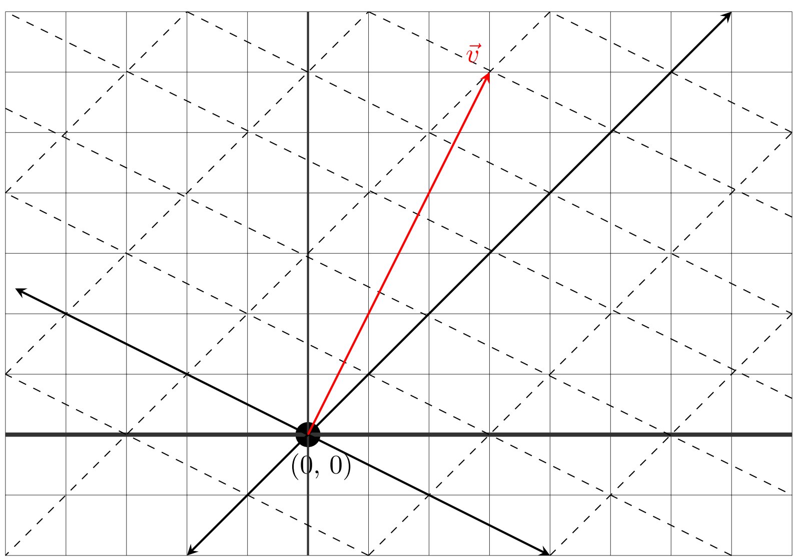
Activity 9.3. The following graph includes the standard coordinate grid defined by the standard basis $\mathcal{E}$ (drawn with solid lines) and the "nonstandard" coordinate grid defined by the basis $\mathcal{B}$ (drawn with dashed lines) from the previous activities. Draw the vectors $\vec{u}, \vec{v}, \vec{w}$ and $\vec{z}$ from the previous activities on the graph below. What do you notice?

Solution.



Activity 9.4. Consider the vector $\vec{v} = 3\vec{e}_1 + 6\vec{e}_2$. Use the graph from the previous activity to find real numbers x_1, x_2 so that $\vec{v} = x_1\vec{b}_1 + x_2\vec{b}_2$.

Solution. Using the grid from Activity 9.3, we can see that $\vec{v} = 5\vec{b}_1 + 1\vec{b}_2$.



Activity 9.5. Let $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \vec{b}_3\}$ be the ordered basis for $\mathbb{R}^3$ where

$$\vec{b}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \vec{b}_2 = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}, \text{ and } \vec{b}_3 = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}.$$

Find the $\mathcal{B}$-coordinates of the vector $\vec{v} = \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix}$.

Solution. We want to find the scalars $c_1, c_2, c_3 \in \mathbb{R}$ such that $\vec{v} = c_1\vec{b_1} + c_2\vec{b}_2 + c_3\vec{b}_3$. This is a vector equation which we can express as a system of linear equations with augmented matrix

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

Row reducing, we find that the reduced row echelon form of this matrix is

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

This gives us a unique solution, where $c_1 = 3, c_2 = -1$, and $c_3 = 1$. In other words,

$$\vec{v}_{\mathcal{B}} = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}.$$

Activity 9.6. Consider the basis $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}$ for $\mathbb{R}^2$ where

$$\vec{b}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \text{ and } \vec{b}_2 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}.$$

(1) Find $M_{\mathcal{B}\leftarrow\mathcal{E}}$ and $M_{\mathcal{E}\leftarrow\mathcal{B}}$.

(2) Show that $M^{-1}_{\mathcal{E}\leftarrow\mathcal{B}} = M_{\mathcal{B}\leftarrow\mathcal{E}}$.

Solution.

(1) Let $M_{\mathcal{B}\leftarrow\mathcal{E}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We know that $M_{\mathcal{B}\leftarrow\mathcal{E}}[\vec{b}_1]_{\mathcal{E}} = [\vec{b}_1]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. So we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

This gives us the following system of linear equations

$$2a + b = 1$$

$$2c + d = 0.$$

Similarly, $M_{\mathcal{B}\leftarrow\mathcal{E}}[\vec{b}_2]_{\mathcal{E}} = [\vec{b}_2]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. So we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 5 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

This gives us the following system of linear equations

$$5a + 3b = 0$$

$$5c + 3d = 1.$$

Combining all four equations and solving them, we get $a = 3, b = -5, c = -1$ and $d = 2$. So that $M_{\mathcal{B}\leftarrow\mathcal{E}} = \begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix}$.

Now to find $M_{\mathcal{E}\leftarrow\mathcal{B}}$, let $M_{\mathcal{E}\leftarrow\mathcal{B}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Note that $M_{\mathcal{E}\leftarrow\mathcal{B}}[\vec{b}_1]_{\mathcal{B}} = [\vec{b}_1]_{\mathcal{E}}$. So we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

And this gives us $a = 2$, and $c = 1$. Similarly, $M_{\mathcal{E}\leftarrow\mathcal{B}}[\vec{b}_2]_{\mathcal{B}} = [\vec{b}_2]_{\mathcal{E}}$. So we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}.$$

And this gives us $b = 5$, and $d = 3$. So, $M_{\mathcal{E}\leftarrow\mathcal{B}} = \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}$.

(2) To show that the two matrices are each others inverses, we only need to show that their product is equal to the identity matrix. This is, in fact, true as we can calculate it directly,

$$M_{\mathcal{E}\leftarrow\mathcal{B}}M_{\mathcal{B}\leftarrow\mathcal{E}} = \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2.$$

Activity 9.7. Let V be the plane in $\mathbb{R}^3$ spanned by

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

(1) Show that $\mathcal{B}$ is also a basis for V, where

$$\mathcal{B} = \left\{ \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix} \right\}.$$

(2) Find the change of basis matrices $M_{\mathcal{C}\leftarrow\mathcal{B}}$ and $M_{\mathcal{C}\leftarrow\mathcal{B}}$.

(3) Find a new basis $\mathcal{D}$ for V that's not equal to $\mathcal{B}$ or $\mathcal{C}$.

Solution.

(1) First let's find $[\vec{b}_1]_{\mathcal{C}}$ and $[\vec{b}_2]_{\mathcal{C}}$. We need to find solutions to the following vector equations

$$\vec{b}_1 = x_1\vec{c}_1 + x_2\vec{c}_2,$$
$$\vec{b}_2 = y_2\vec{c}_1 + y_2\vec{c}_2.$$

Expressing these as systems of linear equations and solving them, we get $\vec{b}_1 = \vec{c}_1 + \vec{c}_2$, and $\vec{b}_2 = 2\vec{c}_1 + \vec{c}_2$. So that $[\vec{b}_1]_{\mathcal{C}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and $[\vec{b}_2]_{\mathcal{C}} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

Now $\begin{pmatrix} [\vec{b}_1]_{\mathcal{C}} & [\vec{b}_2]_{\mathcal{C}} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}$ is invertible because its determinant is equal to $1 \cdot 1 - 2 \cdot 1 = -1 \neq 0$. By Theorem 9.8, we can conclude that $\mathcal{B}$ is a basis for V.

(2) By Theorem 9.8,

$$M_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{pmatrix} [\vec{b}_1]_{\mathcal{C}} & [\vec{b}_2]_{\mathcal{C}} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}.$$

Also,

$$M_{\mathcal{B}\leftarrow\mathcal{C}} = M_{\mathcal{C}\leftarrow\mathcal{B}}^{-1} = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix}.$$

(3) Let $\mathcal{D} = \{\vec{d}_1, \vec{d}_2\}$. Then $\mathcal{D}$ is a basis for V if $\begin{pmatrix} [\vec{d}_1]_\mathcal{C} & [\vec{d}_2]_\mathcal{C} \end{pmatrix}$ is invertible. Let $\begin{pmatrix} [\vec{d}_1]_\mathcal{C} & [\vec{d}_2]_\mathcal{C} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 0 \end{pmatrix}$, the determinant of this matrix is equal to $1 \cdot 0 - 2 \cdot 3 = -6 \neq 0$, and so the matrix is invertible. This matrix gives us the vectors $\vec{d}_1$ and $\vec{d}_2$,

$$\vec{d}_1 = \vec{c}_1 + 3\vec{c}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 3\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 1 \\ 4 \end{pmatrix},$$

$$\vec{d}_2 = 2\vec{c}_1 + 0\vec{c}_2 = 2\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 0\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}.$$

B.10. Chapter 10 Activity Solutions

Activity 10.1. Let $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}$ be the basis with

$$\vec{b}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \vec{b}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

(1) Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 2x + y \\ x - y \end{pmatrix}.$$

Find the defining matrices A_F and $A_{F,\mathcal{B}}$.

(2) Let $G : \mathbb{R}^2 \to \mathbb{R}^2$ be the function which stretches vectors in the $\vec{b}_1$ direction by 2 and leaves vectors in the $\vec{b}_2$ direction fixed. That is,

$$G(x_1\vec{b}_1 + x_2\vec{b}_2) = 2x_1\vec{b}_1 + x_2\vec{b}_2.$$

Find the defining matrices A_G and $A_{G,\mathcal{B}}$.

Solution.

(1) By Theorem 5.4, we have

$$A_F = \begin{pmatrix} F(\vec{e}_1) & F(\vec{e}_2) \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}.$$

By Theorem 10.1, we have $A_{F,\mathcal{B}} = \begin{pmatrix} [F(\vec{b}_1)]_{\mathcal{B}} & [F(\vec{b}_2)]_{\mathcal{B}} \end{pmatrix}$. Now, $[F(\vec{b}_1)]_{\mathcal{E}} = F(\begin{pmatrix} 1 \\ 1 \end{pmatrix}) = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$. We need to find $[F(\vec{b}_1)]_{\mathcal{B}}$. Note that $\begin{pmatrix} 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \vec{b}_1 - \vec{b}_2$. And so $[F(\vec{b}_1)]_{\mathcal{B}} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Next, we find $[F(\vec{b}_2)]_{\mathcal{B}}$. First, $[F(\vec{b}_2)]_{\mathcal{E}} = F(\begin{pmatrix} -2 \\ 1 \end{pmatrix}) = \begin{pmatrix} -3 \\ -3 \end{pmatrix}$. Note that $\begin{pmatrix} -3 \\ -3 \end{pmatrix} = -3\begin{pmatrix} 1 \\ 1 \end{pmatrix} = -3\vec{b}_1 + 0\vec{b}_2$. And so, $[F(\vec{b}_2)]_{\mathcal{B}} = \begin{pmatrix} -3 \\ 0 \end{pmatrix}$. Finally, we have

$$A_{F,\mathcal{B}} = \begin{pmatrix} 1 & -3 \\ -1 & 0 \end{pmatrix}.$$

(2) By Theorem 10.1, we have

$$A_{G,\mathcal{B}} = \begin{pmatrix} [G(\vec{b}_1)]_{\mathcal{B}} & [G(\vec{b}_2)]_{\mathcal{B}} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

By Theorem 5.4, we have $A_G = \begin{pmatrix} [G(\vec{e}_1)]_{\mathcal{E}} & [G(\vec{e}_2)]_{\mathcal{E}} \end{pmatrix}$. In order to apply the function on $\vec{e}_1$ and $\vec{e}_2$, we need to find their coordinated in the basis $\mathcal{B}$. Note that

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{3}\begin{pmatrix} -2 \\ 1 \end{pmatrix} = \frac{1}{3}\vec{b}_1 - \frac{1}{3}\vec{b}_2,$$

$$\vec{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{2}{3}\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} -2 \\ 1 \end{pmatrix} = \frac{2}{3}\vec{b}_1 + \frac{1}{3}\vec{b}_2.$$

Applying G, we get

$$G([\vec{e}_1]_{\mathcal{B}}) = G(\frac{1}{3}\vec{b}_1 - \frac{1}{3}\vec{b}_2) = \frac{2}{3}\vec{b}_1 - \frac{1}{3}\vec{b}_2,$$
$$G([\vec{e}_2]_{\mathcal{B}}) = G(\frac{2}{3}\vec{b}_1 + \frac{1}{3}\vec{b}_2) = \frac{4}{3}\vec{b}_1 + \frac{1}{3}\vec{b}_2.$$

Finally,

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

Activity ??. Let $B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Find a matrix C that's similar to B, but not equal to B.

Solution. Suppose B is the defining matrix for the transformation F in the standard basis. Let $\mathcal{C} = \{\vec{c}_1, \vec{c}_2\}$, where $\vec{c}_1 = \vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and $\vec{c}_2 = -\vec{e}_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$. We will find $A_{F,\mathcal{C}}$. By Theorem 10.1, $A_{F,\mathcal{C}} = \begin{pmatrix} [F(\vec{c}_1)]_{\mathcal{C}} & [F(\vec{c}_2)]_{\mathcal{C}} \end{pmatrix}$.Then $F(\vec{c}_1) = F(\vec{e}_1) = \vec{e}_1 + 3\vec{e}_2 = \vec{c}_1 - 3\vec{c}_2$, and $[F(\vec{c}_1)]_{\mathcal{C}} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$. Also, $F(\vec{c}_2) = F(-\vec{e}_2) = -F(\vec{e}_2) = -(2\vec{e}_1 + 4\vec{e}_2) = -2\vec{e}_1 - 4\vec{e}_2 = -2\vec{c}_1 + 4\vec{c}_2$, and $[F(\vec{c}_2)]_{\mathcal{C}} = \begin{pmatrix} -2 \\ 4 \end{pmatrix}$. Let $C = \begin{pmatrix} 1 & -2 \\ -3 & 4 \end{pmatrix}$. Then $B = A_{F,\mathcal{E}}$, and $C = A_{F,\mathcal{C}}$, and so they are similar.

Activity ??. Show that if two matrices B and C are similar, then B is invertible if and only if C is invertible.

Solution. If two matrices B and C are similar, then there exists an invertible matrix P such that $B = PCP^{-1}$. Note that $\det(B) = \det(PCP^{-1}) = \det(P)\det(C)\det(P^{-1}) = \det(P)\det(C)\frac{1}{\det(P)} = \det(C)$. And so $\det(B) = 0$ if and only if $\det(C) = 0$. It follows from Theorem 7.18 that B is invertible if and only if C is invertible.

Activity ??. Show that the matrices

$$B = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \text{ and } C = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 1 & -1 & -1 \end{pmatrix}$$

are not similar.

Solution. Note that $\det(B) = 2$ and so B is invertible, while $\det(C) = 0$ and so C is not invertible. By the previous activity, if B and C are similar, they are either both invertible or both non-invertible. It follows that B and C are not similar matrices.

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

(1) Show that A has eigenvalues $\lambda = -1$ and $\lambda = 4$.
(2) Find vectors $\vec{v}$ and $\vec{w}$ so that $E_{-1} = \operatorname{Span}(\vec{v})$ and $E_4 = \operatorname{Span}(\vec{w})$.
(3) Observing that $\{\vec{v}, \vec{w}\}$ forms a basis for $\mathbb{R}^2$, show that A is similar to the matrix $\begin{pmatrix} -1 & 0 \\ 0 & 4 \end{pmatrix}$.
(4) Give a geometric description for how the function T_A transforms $\mathbb{R}^2$.

Solution.

(1) We begin by finding the characteristic polynomial of A,

$$\chi_A(x) = \det(A - xI) = \det\begin{pmatrix} 1-x & 2 \\ 3 & 2-x \end{pmatrix} = (1-x)(2-x)-6 = 2-3x+x^2-6 = x^2-3x-4 = (x-4)(x+1).$$

The eigenvalues of A are the roots of $\chi_A(x)$, and so $\lambda_1 = 4$, and $\lambda_2 = -1$.

(2) The λ-eigenspace E_λ is equal to the null space of $A - \lambda I$. For $\lambda_1 = 4$, we have $A - 4I = \begin{pmatrix} -3 & 2 \\ 3 & -2 \end{pmatrix}$ which has RREF equal to $\begin{pmatrix} 1 & \frac{-2}{3} \\ 0 & 0 \end{pmatrix}$. This has null space equal to the span of $\vec{w} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$. And $E_4 = \text{Span}(\vec{w})$. Now to find the eigenspace of $\lambda_2 = -1$, we have $A + \lambda I = \begin{pmatrix} 2 & 2 \\ 3 & 3 \end{pmatrix}$, which has RREF equal to $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$. This has null space equal to the span of $\vec{v} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. And $E_{-1} = \text{Span}(\vec{v})$.

(3) Let A be the defining matrix for a transformation F on $\mathbb{R}^2$ in the standard basis, so that $A = A_F$. Let $\mathcal{B} = \{\vec{v}, \vec{w}\}$, we will find $A_{F,\mathcal{B}}$. Note that

$$F(\vec{v}) = A\vec{v} = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -\vec{v} + 0\vec{w}$$

. Also,

$$F(\vec{w}) = A\vec{w} = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 8 \\ 12 \end{pmatrix} = 0\vec{v} + 4\vec{w}.$$

By Theorem 10.1, we have $A_{F,\mathcal{B}} = \begin{pmatrix} -1 & 0 \\ 0 & 4 \end{pmatrix}$. It follows that the matrices $\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$ and $\begin{pmatrix} -1 & 0 \\ 0 & 4 \end{pmatrix}$ are similar.

(4) The function T_A stretches vectors in the direction of $\vec{v}$ by a factor of -1 and stretches vectors in the direction of $\vec{w}$ by a factor of 4.

Activity ??. Consider the matrix $B = \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix}$.

(1) Show that B has exactly one eigenvalue $\lambda = 2$, and that the geometric multiplicity of $\lambda = 2$ is equal to 1.

(2) Do you think it's possible for B to be similar to a matrix of the form

$$\begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}?$$

Explain why or why not.

Solution.

(1) First, we find the characteristic polynomial of B,

$$\chi_B(x) = \det(B - xI) = \det\begin{pmatrix} 2-x & 2 \\ 0 & 2-x \end{pmatrix} = (2-x)^2.$$

The eigenvalues of B are the roots of the characteristic polynomial, and so we have only one eigenvalue $\lambda = 2$ with algebraic multiplicity equal to 2. Next, we find 2-eigenspace of B, which is the null space of $B - 2I$. We have $B - 2I = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$, and this has null space equal to Span $\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}$. Since the null space (the eigenspace) is 1-dimensional, the geometric multiplicity of $\lambda = 2$ is equal to 1.

(2) It is not. Let F be the transformation with defining matrix B, so $A_F = B$. There will be no basis $\mathcal{C}$ with $A_{F,\mathcal{C}} = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}$. Note that C has an associated transformation that stretches the first component of a vector by a factor of d_1, and the second component by a factor of d_2. But our transformation F only has one eigenvalue with one eigenvector, so only vectors in the direction of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ are stretched by a factor of 2. We do not have a second (linearly independent) vector that can be stretched by any other eigenvalue.

Activity ??. Determine which of the following matrices A_i are diagonalizable. For those that are, find an invertible matrix C_i and diagonal matrix D_i so that $D_i = C_i^{-1} A_i C_i$.

$$A_1 = \begin{pmatrix} 0 & 1 \\ 2 & 1 \end{pmatrix} \; A_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{pmatrix} \; A_3 = \begin{pmatrix} 4 & -3 & 0 \\ 2 & -1 & 0 \\ 1 & -1 & 1 \end{pmatrix}.$$

Solution. Note that A_1 has eigenvalues 2 and -1. Also,

$$E_2 = \text{Span}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right)$$

$$E_{-1} = \text{Span}\left(\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right).$$

Observe that

$$\vec{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ and } \vec{w} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

are linearly independent eigenvectors, and so by the Diagonalization Theorem, the matrix A_1 is diagonalizable. If we let

$$D_1 = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \text{ and } C_1 = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}$$

then we can see that $D_1 = C_1^{-1} A_1 C_1$.

Note that A_2 has eigenvalues 2 and 1. Also,

$$E_2 = \text{Span}\left(\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}\right)$$

$$E_1 = \text{Span}\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right).$$

Since the eigenvectors in E_2 (resp E_1) are linearly dependent, it's not possible to have three linearly independent eigenvectors. So, by the Diagonalization Theorem, we know that A_2 is not diagonalizable.

Note that A_3 has eigenvalues 2 and 1. Also,

$$E_2 = \text{Span}\left(\begin{pmatrix}3\\2\\1\end{pmatrix}\right)$$

$$E_1 = \text{Span}\left(\begin{pmatrix}0\\0\\1\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}\right).$$

Since the vectors

$$\vec{v}_1 = \begin{pmatrix}3\\2\\1\end{pmatrix}, \vec{v}_2 = \begin{pmatrix}0\\0\\1\end{pmatrix}, \text{ and } \vec{v}_3 \begin{pmatrix}1\\1\\0\end{pmatrix}$$

are linearly independent, then by the Diagonalization theorem A_3 is Diagonalizable. If we let

$$D_3 = \begin{pmatrix}2 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{pmatrix} \text{ and } C_3 = \begin{pmatrix}3 & 0 & 1\\2 & 0 & 1\\1 & 1 & 0\end{pmatrix}$$

then we can see that $D_3 = C_3^{-1} A_3 C_3$.

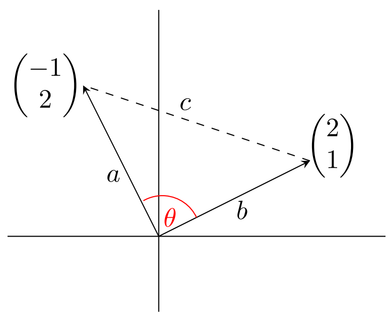
B.11. Chapter 11 Activity Solutions

Activity 11.1. Use geometric reasoning to show that the vectors $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ are perpendicular to each other.

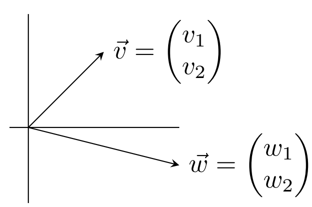
Solution. Consider the vectors plotted in the graph below. Recall the law of cosines,

$$c^2 = a^2 + b^2 - 2ab\cos\theta.$$

Using the Pythagorean Theorem, we can see that the lengths of the vectors are $a = b = \sqrt{(\pm 1)^2 + 2^2} = \sqrt{5}$. We also need to the distance between the two vectors, which we can also find using the Pythagorean Theorem by looking at the right angled triangle with vertices $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$, $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. This will give us $c = \sqrt{(2 - -1)^2 + (2-1)^2} = \sqrt{10}$. Applying the law of cosines, we find that $\cos\theta = 0$ (note that $a^2 + b^2 = 5 + 5 = 10 = c^2$ and so $-2ab\cos\theta = 0$, but since a and b are non-zero, it follows that $\cos\theta = 0$). Therefore, it must be the case that $\theta = 90°$.



Activity 11.2. Let $\vec{v}$ and $\vec{w}$ be the vectors drawn below.



Find a formula for the length of $\vec{v}$ and $\vec{w}$ and the distance between $\vec{v}$ and $\vec{w}$.

Solution. We can find a formula for the length of a vector using the Pythagorean Theorem. We have that the length of a vector $\vec{v}$ is equal to $\sqrt{v_1^2 + v_2^2} = \sqrt{v_1 v_1 + v_2 v_2}$. Similarly, the length of the vector $\vec{w}$ is equal to $\sqrt{w_1^2 + w_2^2}$. Now note that the distance between the two vectors is just the length of the vector $\vec{v} - \vec{w}$, and so the distance between $\vec{v}$ and $\vec{w}$ is equal to $\sqrt{(v_1 - w_1)^2 + (v_2 - w_2)^2}$.

Activity 11.3. Let $\vec{u}$ and $\vec{v}$ be any two vectors in $\mathbb{R}^2$. In this activity, we'll derive a formula for the angle between $\vec{u}$ and $\vec{v}$.

(1) Show that for any vector $\vec{x}$ in $\mathbb{R}^n$ we have $\|\vec{x}\|^2 = \vec{x} \cdot \vec{x}$.

(2) Use law of cosines to show that the angle between $\vec{u}$ and $\vec{v}$ is given by

$$\theta = \arccos \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|\|\vec{v}\|}.$$

(Hint: use the observation in part (1), along with Proposition 11.2)

(3) Conclude that $\vec{u}$ and $\vec{v}$ are perpendicular if and only if $\vec{u} \cdot \vec{v} = 0$.

Solution.

(1) Let $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$. By definition, $\|\vec{x}\| = \sqrt{\vec{x} \cdot \vec{x}}$, squaring both sides, we get $\|\vec{x}\|^2 = \vec{x} \cdot \vec{x}$.

(2) The law of cosines tell us that

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\| \cos\theta.$$

And this gives us the following,

$$\begin{aligned}
\cos\theta &= \frac{1}{2\|\vec{u}\|\|\vec{v}\|}(\|\vec{u}\|^2 + \|\vec{v}\|^2 - \|\vec{u} - \vec{v}\|^2) \\
&= \frac{1}{2\|\vec{u}\|\|\vec{v}\|}(\vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - (\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v})) \\
&= \frac{1}{2\|\vec{u}\|\|\vec{v}\|}(\vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - [(\vec{u} - \vec{v}) \cdot \vec{u} - (\vec{u} - \vec{v}) \cdot \vec{v})]) && \text{using Proposition 11.2} \\
&= \frac{1}{2\|\vec{u}\|\|\vec{v}\|}(\vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - [\vec{u} \cdot \vec{u} - \vec{v} \cdot \vec{u} - \vec{u} \cdot \vec{v} + \vec{v} \cdot \vec{v}]) && \text{using Proposition 11.2} \\
&= \frac{1}{2\|\vec{u}\|\|\vec{v}\|}(\vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{v} - \vec{v} \cdot \vec{v}) && \text{using Proposition 11.2} \\
&= \frac{1}{2\|\vec{u}\|\|\vec{v}\|} 2\vec{u} \cdot \vec{v} \\
&= \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|\|\vec{v}\|}.
\end{aligned}$$

Therefore, $\theta = \arccos \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|\|\vec{v}\|}$ as required.

(3) Two vectors are perpendicular if and only if θ, the angle between them, is equal to $90°$. Since $\cos 90° = 0$, it follows that the vectors are perpendicular if and only if $\frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|\|\vec{v}\|} = 0$. Multiplying both sides of the last equation by $\|\vec{u}\|\|\vec{v}\|$ gives us $\vec{u} \cdot \vec{v} = 0$. It follows that the vectors $\vec{u}$ and $\vec{v}$ are perpendicular if and only if $\vec{u} \cdot \vec{v} = 0$.

Activity ??. Determine which of the following pairs of vectors $\vec{u}$ and $\vec{v}$ are orthogonal.

(1) $\vec{u} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$

(2) $\vec{u} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$

(3) $\vec{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$

Solution. For each pair of vectors, we need only check the dot product of the two vectors.

(1) Note that

$$\vec{u} \cdot \vec{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 2 \end{pmatrix} = (2)(-1) + (1)(2) = 0.$$

And so the vectors are orthogonal.

(2) Note that

$$\vec{u} \cdot \vec{v} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = (-1)(2) + (0)(1) + (1)(1) = -1 \neq 0.$$

It follows that the two vectors are NOT orthogonal.

(3) Note that

$$\vec{u} \cdot \vec{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix} = (1)(1) + (2)(-1) + (3)(0) + (1)(1) = 0.$$

It follows that the vectors are orthogonal.

Activity ??. Consider the bases $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ for $\mathbb{R}^2$ given by

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}, \mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \text{ and } \mathcal{D} = \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right\},$$

(1) Determine which of the bases above are orthogonal and which are orthonormal.

(2) Calculate $\vec{u} \cdot \vec{u}$ given that $[\vec{u}]_{\mathcal{B}} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$.

(3) Calculate $\vec{v} \cdot \vec{v}$ given that $[\vec{v}]_{\mathcal{C}} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$.

(4) Calculate $\vec{w} \cdot \vec{w}$ given that $[\vec{w}]_{\mathcal{D}} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$.

(5) What did you notice in your calculations?

Solution.

(1) For basis $\mathcal{B}$, note that $\begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \neq 0$. Since the basis vectors are not orthogonal, the basis $\mathcal{B}$ is not orthogonal either.

For basis $\mathcal{C}$, note that $\begin{pmatrix}1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\-1\end{pmatrix} = 0$. Since the basis vectors are orthogonal, the basis $\mathcal{C}$ is also orthogonal. Now the vectors in the basis $\mathcal{C}$ have norm equal to $\sqrt{2}\neq 1$, and so the basis is not orthonormal.

For basis $\mathcal{D}$, note that $\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}\cdot\begin{pmatrix}\frac{1}{\sqrt{2}}\\-\frac{1}{\sqrt{2}}\end{pmatrix} = 0$. Since the basis vectors are orthogonal, the basis $\mathcal{D}$ is also orthogonal. Also note that

$$\left\|\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}\right\| = \left\|\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{-1}{\sqrt{2}}\end{pmatrix}\right\| = \sqrt{(\frac{1}{\sqrt{2}})^2+(\frac{\pm1}{\sqrt{2}})^2} = \sqrt{\frac{1}{2}+\frac{1}{2}} = \sqrt{1} = 1.$$

It follows that the basis $\mathcal{D}$ is orthonormal.

(2) We have

$$\begin{aligned}\vec{u}\cdot\vec{u} &= \left(u_1\begin{pmatrix}1\\2\end{pmatrix}+u_2\begin{pmatrix}1\\1\end{pmatrix}\right)\cdot\left(u_1\begin{pmatrix}1\\2\end{pmatrix}+u_2\begin{pmatrix}1\\1\end{pmatrix}\right)\\ &= u_1^2\begin{pmatrix}1\\2\end{pmatrix}\cdot\begin{pmatrix}1\\2\end{pmatrix}+u_1u_2\begin{pmatrix}1\\2\end{pmatrix}\cdot\begin{pmatrix}1\\1\end{pmatrix}+u_2u_1\begin{pmatrix}1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\2\end{pmatrix}+u_2^2\begin{pmatrix}1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\1\end{pmatrix}\\ &= 5u_1^2+3u_1u_2+3u_1u_2+2u_2^2\\ &= 5u_1^2+6u_1u_2+2u_2^2.\end{aligned}$$

(3) We have

$$\begin{aligned}\vec{v}\cdot\vec{v} &= \left(v_1\begin{pmatrix}1\\1\end{pmatrix}+v_2\begin{pmatrix}1\\-1\end{pmatrix}\right)\cdot\left(v_1\begin{pmatrix}1\\1\end{pmatrix}+v_2\begin{pmatrix}1\\-1\end{pmatrix}\right)\\ &= v_1^2\begin{pmatrix}1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\1\end{pmatrix}+v_1v_2\begin{pmatrix}1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\-1\end{pmatrix}+v_2v_1\begin{pmatrix}1\\-1\end{pmatrix}\cdot\begin{pmatrix}1\\1\end{pmatrix}+v_2^2\begin{pmatrix}1\\-1\end{pmatrix}\cdot\begin{pmatrix}1\\-1\end{pmatrix}\\ &= 2v_1^2+2v_2^2.\end{aligned}$$

(4) We have

$$\begin{aligned}\vec{w}\cdot\vec{w} &= \left(w_1\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}+w_2\begin{pmatrix}\frac{1}{\sqrt{2}}\\-\frac{1}{\sqrt{2}}\end{pmatrix}\right)\cdot\left(w_1\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}+w_2\begin{pmatrix}\frac{1}{\sqrt{2}}\\-\frac{1}{\sqrt{2}}\end{pmatrix}\right)\\ &= w_1^2\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}\cdot\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}+w_1w_2\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}\cdot\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{-1}{\sqrt{2}}\end{pmatrix}+w_2w_1\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{-1}{\sqrt{2}}\end{pmatrix}\cdot\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\end{pmatrix}+w_2^2\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{-1}{\sqrt{2}}\end{pmatrix}\cdot\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{-1}{\sqrt{2}}\end{pmatrix}\\ &= w_1^2+w_2^2.\end{aligned}$$

(5) We notice that for an orthogonal basis with vectors that have the same length, the length of a vector is the same as the dot product of the vector with itself scaled by the length of the basis vectors. And for an orthonormal basis, the length of the vector is the same as the dot product of the vector by itself!

Activity ??. Suppose that $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}$ is an orthonormal basis for $\mathbb{R}^2$, and consider the matrix $A = \begin{pmatrix}\vec{b}_1 & \vec{b}_2\end{pmatrix}$. Show that A is invertible with $A^{-1} = A^\top$.

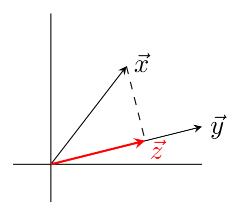
Solution. Let $\vec{b}_1 = \begin{pmatrix} b_{11} \\ b_{21} \end{pmatrix}, \vec{b}_2 = \begin{pmatrix} b_{12} \\ b_{22} \end{pmatrix}$. Then $A = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$, and $A^\top = \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix}$. We have

$$\begin{aligned} A^\top A &= \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \\ &= \begin{pmatrix} \vec{b}_1^\top \vec{b}_1 & \vec{b}_1^\top \vec{b}_2 \\ \vec{b}_2^\top \vec{b}_1 & \vec{b}_2^\top \vec{b}_2 \end{pmatrix} \\ &= \begin{pmatrix} \vec{b}_1 \cdot \vec{b}_1 & \vec{b}_1 \cdot \vec{b}_2 \\ \vec{b}_2 \cdot \vec{b}_1 & \vec{b}_2 \cdot \vec{b}_2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

It follows that A is invertible and $A^{-1} = A^\top$.

Activity 11.7. Let $\vec{x}$ and $\vec{y}$ be vectors in $\mathbb{R}^2$, and let $\vec{z}$ be the closest point in $\mathrm{Span}(\vec{y})$ to $\vec{x}$.

(1) Use the picture below to argue that $\vec{y}$ is orthogonal to $\vec{x} - \vec{z}$.



(2) Since $\vec{z}$ is in $\mathrm{Span}(\vec{y})$, we can write $\vec{z} = c\vec{y}$ for some real number c. Use the previous part to show that

$$c = \frac{\vec{x} \cdot \vec{y}}{\vec{y} \cdot \vec{y}}.$$

(3) Conclude that the closest point on $\mathrm{Span}(\vec{y})$ to $\vec{x}$ is given by

$$\vec{z} = \frac{\vec{x} \cdot \vec{y}}{\vec{y} \cdot \vec{y}} \vec{y}.$$

Solution.

(1) The dashed line gives us the vector $\vec{x} - \vec{z}$, which we can root at the origin. We want to show that the dashed line makes a $90°$ angle with the vector $\vec{y}$. Suppose it does not, then the perpendicular line dropped from $\vec{x}$ to the line $\mathrm{Span}(\vec{y})$ will be shorter than the dashed line (by the Pythagorean Theorem, since the dashed line is the hypotenuse of a right-angled triangle with one of the sides being the perpendicular line dropped from $\vec{x}$). But this contradicts that $\vec{z}$ is the closest point in $\mathrm{Span}(\vec{y})$ to $\vec{x}$. It follows that the dashed line, or the vector $\vec{x} - \vec{z}$ is orthogonal to $\vec{y}$.

(2) From the previous part, we know that $\vec{y} \cdot (\vec{x} - \vec{z}) = 0$. But this means that $\vec{y} \cdot (\vec{x} - c\vec{y}) = 0$, and we have

$$\begin{aligned}\vec{y} \cdot (\vec{x} - c\vec{y}) &= 0\\ \vec{y} \cdot \vec{x} - c\vec{y} \cdot \vec{y} &= 0\\ \vec{x} \cdot \vec{y} &= c\vec{y} \cdot \vec{y}\\ c &= \frac{\vec{x} \cdot \vec{y}}{\vec{y} \cdot \vec{y}}.\end{aligned}$$

(3) Thus the closest point $\vec{z}$ on $\operatorname{Span}(\vec{y})$ to $\vec{x}$ is given by

$$\begin{aligned}\vec{z} &= c\vec{y}\\ &= \frac{\vec{x} \cdot \vec{y}}{\vec{y} \cdot \vec{y}}\vec{y}.\end{aligned}$$

Activity 11.8. Consider the matrix $A = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}$.

(1) Check that this matrix is symmetric.

(2) Verify that $\chi_A(x) = (7 - x)(2 - x)$ and that

$$E_7 = \operatorname{Span}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right) \text{ and } E_2 = \operatorname{Span}\left(\begin{pmatrix} -2 \\ 1 \end{pmatrix}\right).$$

(3) Find an ortho*normal* basis of eigenvectors for A.

(4) Find a spectral deomposition for A. That is, find an orthogonal matrix Q and diagonal matrix A so that $A = QDQ^{\top}$.

(5) Since Q is a rotation matrix, we can write

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Find the value of θ satisfying the equality above.

(6) Using your work in the previous parts, give a complete geometric description of how T_A transforms $\mathbb{R}^2$.

Solution.

(1) Note that $A^{\top} = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} = A$. Therefore, A is symmetric.

(2) First we find $\chi_A(x)$,

$$\begin{aligned}\chi_A(x) &= \det(A - xI)\\ &= \det\begin{pmatrix} 3 - x & 2 \\ 2 & 6 - x \end{pmatrix}\\ &= (3 - x)(6 - x) - 4\\ &= 14 - 9x + x^2\\ &= (7 - x)(2 - x).\end{aligned}$$

Next we find the eigenspaces, E_7 is the null space of $A - 7I = \begin{pmatrix} -4 & 2 \\ 2 & -1 \end{pmatrix}$, and so $E_7 = \text{Span}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right)$. And E_2 is the null space of $A - 2I = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$, and so $E_2 = \text{Span}\left(\begin{pmatrix} -2 \\ 1 \end{pmatrix}\right)$.

(3) Note that $\left\{\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \end{pmatrix}\right\}$ is an orthogonal basis for $\mathbb{R}^2$ of eigenvectors for A. So we only need to normalize it. The set $\left\{\begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}, \begin{pmatrix} \frac{-2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}\right\}$ is an orthonormal basis for $\mathbb{R}^2$ of eigenvectors for A.

(4) By the Diagonalization Theorem, we have

$$A = QDQ^{-1} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 7 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}^{-1}.$$

Since, Q is an orthogonal matrix, we have $Q^{-1} = Q^\top$, so that

$$A = QDQ^{-1} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 7 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}^{\top}.$$

This gives a spectral decomposition of A.

(5) We can write

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}.$$

Solving for θ, we have $\cos\theta = \frac{1}{\sqrt{5}}$, and this gives $\theta = \arctan(\frac{1}{\sqrt{5}}) \approx 63°$.

(6) We have $T_A = T_{QDQ^\top} = T_Q T_D T_{Q^T}$. So T_A rotates a vector in $\mathbb{R}^2$ clockwise by 63°, then it dilates it by scaling its x-component by a factor of 7 and its y-component by a factor of 2, and finally it rotates the vector 63° counter-clockwise.

Activity ??. Find the singular values of the following matrices. Given your calculations, what can you say about the corresponding transformations?

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

(2) $B = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{pmatrix}$

(3) $C = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

Solution.

(1) The singular values of A are the square roots of the eigenvalues of

$$A^\top A = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 4 \end{pmatrix}.$$

Since the eigenvalues of $A^\top A$ are 4 and 5, the singular values of A are 2 and $\sqrt{5}$.

(2) The singular values of B are the square roots of the eigenvalues of

$$B^\top B = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 3 \\ 3 & 6 & 3 \\ 3 & 3 & 6 \end{pmatrix}.$$

Now $\chi_{B^\top B}(x) = -x^3 + 14x^2 - 33x = x(x-3)(x-11)$. Since the eigenvalues of $B^\top B$ are 0, 3 and 11, the singular values of B are 0, $\sqrt{3}$ and $\sqrt{11}$.

(3) The singular values of C are the square roots of the eigenvalues of

$$C^\top C = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Since $C^\top C$ has only one eigenvalue equal to 2, the singular values of C are $\sqrt{2}$.

Appendix C

Chapter Exercise Solutions

C.1. Chapter 1 Exercise Solutions

P1.1 Note that an $m \times n$ matrix A cannot have more than one pivot per row. This means that it can have at most m pivots. The matrix A also cannot have more than one pivot per column, meaning that it can have at most n pivots. Combining these findings, we see that A can have at most $\min\{m, n\}$ pivots.

P1.2 We're considering matrices with 3 rows and 4 columns that have exactly 2 pivots. This leaves us with one row that does not have a pivot, so we must have a row of zeros. But in RREF a row of zeros must be at the bottom of the matrix, so the third row must be a row of zeros, and the two pivots must be in the first two rows. We'll then need to figure out which columns the pivots should be in. The matrices in RREF with a pivot in the first column are:

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

The 3×4 matrices in RREF with exactly 2 pivots in the first two rows, no pivot in the first column, and a pivot in the second column are:

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

The 3×4 matrix in RREF with exactly 2 pivots in the first two rows, and no pivots in the first 2 columns is:

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

There are no other matrices with pivots in the first two rows.

P1.3 (a) No, a homogeneous system always has at least one solution, namely the trivial solution (where all variables are set to zero).

(b) Yes, a homogeneous system can have exactly one solution. For example, the following system of equations

$$x + y = 0$$

$$x - y = 0$$

has exactly one solution, when x= y =0 (the trivial solution).

(c) Yes, a homogeneous system can have infinitely many solutions. For example, the following system of equations

$$x - y = 0$$

$$2x - 2y = 0$$

has infinitely many solutions. Any point on the line $y = x$ is a solution to the system of equations.

P1.4 (a) The following system of equations has a unique solution, and the RREF of its augmented matrix has a row of zeros:

$$x + y = 1$$

$$2x + 3y = 4$$

$$3x + 4y = 5$$

(b) There is no system of linear equations with a unique solution that has fewer equations than variables. If we have m equations in n variables and $m < n$, then our coefficient matrix can have at most m pivots in its RREF. So, there will be at least one column with no pivot, a free variable column. If the system is consistent, the existence of a free variable implies that there will be infinitely many solutions to the system of linear equations.

(c) The following is a system of linear equations with a unique solution that has three equations in two variables:

$$x + y = 0$$

$$x - y = 0$$

$$3x + 4y = 0$$

(d) There is no consistent system in two variables whose augmented matrix has exactly three pivot columns. A system of equations with two variables would have an augmented matrix with exactly three columns. If the augmented matrix has three pivot columns, it means that every column of the matrix is a pivot column, in particular, the last column of the RREF of the augmented matrix will have a pivot. By Theorem 1.26, this would mean that the system is inconsistent.

(e) There is no inconsistent system whose coefficient matrix has a pivot in every row. If the coefficient matrix of a system of linear equations has a pivot in every row, then the RREF of the augmented matrix cannot have a pivot in the last column (since there is at most one pivot in every row and the pivots are all already in the coefficient matrix).

P1.5 Suppose we have the system of equations

$$ax + by = r$$

$$cx + dy = s$$

and that $ad - bc \neq 0$. The augmented matrix representing this system of linear equations is given by

$$\left(\begin{array}{cc|c} a & b & r \\ c & d & s \end{array}\right).$$

Let's row reduce this matrix to find out when the system is consistent. There are two cases to consider.

Case 1: If $a = 0$, then the condition $ad - bc \neq 0$ implies that $bc \neq 0$. In this case, we have the augmented matrix

$$\left(\begin{array}{cc|c} 0 & b & r \\ c & d & s \end{array}\right),$$

which has two pivot columns (since b and c are non-zero) in the coefficient matrix and the last column is therefore not a pivot column. Hence the system is consistent with a unique solution.

Case 2: If $a \neq 0$, then we can perform the following row operations,

$$\begin{aligned} \left(\begin{array}{cc|c} a & b & r \\ c & d & s \end{array}\right) &\sim \left(\begin{array}{cc|c} a & b & r \\ 0 & d - \frac{bc}{a} & s - \frac{cr}{a} \end{array}\right), \ R_2 - \tfrac{c}{a}R_1 \\ &\sim \left(\begin{array}{cc|c} a & b & r \\ 0 & ad - bc & as - cr \end{array}\right), \ aR_2. \end{aligned}$$

Since $a \neq 0$ and $ad - bc \neq 0$, we can see that there will be two pivot columns in the coefficient matrix, and the last column will not have a pivot. This means that the system is consistent with a unique solution.

C.2. Chapter 2 Exercise Solutions

P2.1 The statement is true. Suppose we have vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^2$ such that $\operatorname{Span}(\vec{u}, \vec{v}) = \mathbb{R}^2$. Let $\vec{x} \in \mathbb{R}^2$ be an arbitrary vector. Since $\vec{u}$ and $\vec{v}$ span $\mathbb{R}^2$, there exist scalars $a, b \in \mathbb{R}$ such that $\vec{x} = a\vec{u} + b\vec{v}$. But then we have

$$\vec{x} = (a - b)\vec{u} + b(\vec{u} + \vec{v}),$$

which shows that the arbitrary vector $\vec{x}$ is in the span of $\vec{u}$ and $\vec{u} + \vec{v}$. Hence, $\operatorname{Span}(\vec{u}, \vec{u} + \vec{v}) = \mathbb{R}^2$.

P2.2 The statement is false. Consider the vectors

$$\vec{u} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \vec{v} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \vec{w} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Any vector in $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ can be written as $x\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + z\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, and so $\text{Span}(\vec{u}, \vec{v}, \vec{w}) = \mathbb{R}^3$. However, $\text{Span}(\vec{u}+\vec{v}, \vec{v}+\vec{w}) \neq \mathbb{R}^3$ since

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \notin \text{Span}(\vec{u}+\vec{v}, \vec{v}+\vec{w}).$$

To show this, let's assume that the vector

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \in \text{Span}(\vec{u}+\vec{v}, \vec{v}+\vec{w}) = \text{Span}\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right).$$

Then there are scalars $a, b \in \mathbb{R}$ such that $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = a\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$. Translating this into a system of linear equations, we get

$$a + b \cdot 0 = 1$$

$$a + b = 0$$

$$a \cdot 0 + b = 0.$$

The first equation gives us $a = 1$ and the third equation gives us $b = 0$. But the second equation tells us we must have $a + b = 0$, while the other two equations give us $a+b = 1+0 = 1$. This means that our system is inconsistent and has no solution.

P2.3 (a) The following are three linearly independent vectors in $\mathbb{R}^4$: $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$.

You can check their linear independence by looking at the matrix whose columns are the vectors above. The matrix

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

is in RREF and has a pivot in every column. By Rouché-Capelli (Theorem 1.26), the vector equation

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

has only one solution, and so the vectors are linearly independent.

(b) The following are three linearly dependent vectors in $\mathbb{R}^4$: $\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\0\\0\end{pmatrix}$.

Note that the last vector is the sum of the first two, and so the last vector is in the span of the other two vectors.

(c) It is not possible to have three linearly independent vectors in $\mathbb{R}^2$. Let $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^2$ be any arbitrary vectors. Now to test for the linear independence of these vectors, we will solve the vector equation

$$a_1\vec{v}_1 + a_2\vec{v}_2 + a_3\vec{v}_3 = \vec{0}.$$

Note that this vector equation has at least one solution (the trivial solution, where $a_1 = a_2 = a_3 = 0$), so we know the system of equations is consistent. Translating this into an augmented matrix, we get a matrix with 2 rows and 4 columns. This matrix can have at most 2 pivot columns (since there are only 2 rows), so there is at least one column in the coefficient matrix that will not have a pivot. By Rouché-Capelli (Theorem 1.26), this means that the system of equations has infinitely many solutions. By the algebraic definition of linear dependence, we can see that any three vector in $\mathbb{R}^2$ must be linearly dependent.

(d) It is not possible to have four linearly independent vectors in $\mathbb{R}^3$. We follow the same argument as we did for the previous part. Let $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^3$ be any arbitrary vectors. Now to test for the linear independence of these vectors, we will solve the vector equation

$$a_1\vec{v}_1 + a_2\vec{v}_2 + a_3\vec{v}_3 + a_4\vec{v}_4 = \vec{0}.$$

Note that this vector equation has at least one solution (the trivial solution, where $a_1 = a_2 = a_3 = a_4 = 0$), so we know the system of equations is consistent. Translating this into an augmented matrix, we get a matrix with 3 rows and 5 columns. This matrix can have at most 3 pivot columns (since there are only 3 rows), so there is at least one column in the coefficient matrix that will not have a pivot. By Rouché-Capelli (Theorem 1.26), this means that the system of equations has infinitely many solutions. By the algebraic definition of linear dependence, we can see that any four vector in $\mathbb{R}^3$ must be linearly dependent.

(e) The following is a 4×3 matrix with linearly independent columns:

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

We know the columns are linearly independent because the matrix is in RREF and has a pivot in every column and so by Rouché-Capelli (Theorem 1.26), the vector equation

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

has only one solution.

(f) There is no 3×4 matrix with linearly independent columns. This would imply the existence of 4 linearly independent vectors in $\mathbb{R}^3$ which we proved is impossible in a previous part.

P2.4 The statement is true. Let $\vec{v}_1, \ldots, \vec{v}_m$ in $\mathbb{R}^n$, where $m > n$. Consider the vector equation $a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a + m\vec{v}_m = \vec{0}$. Looking at the augmented matrix for this system, we see that it has n rows and $m+1$ columns. Since the last column is all zeros, it cannot be a pivot column, and so the system of equations is consistent. Since there are only n rows, there can be at most n pivot columns. And since the number of columns of the coefficient matrix m is greater than the number of rows, it follows that there is at least one column that will not have a pivot in the RREF of the coefficient matrix. By Rouché-Capelli (Theorem 1.26), the vector equation

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a + m\vec{v}_m = \vec{0}$$

has infinitely many solutions. Then, by the algebraic definition of linear dependence, it follows that the vectors $\vec{v}_1, \ldots, \vec{v}_m$ are linearly dependent.

P2.5 The statement is false. Consider the vectors

$$\vec{v}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}, \vec{v}_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}, \vec{v}_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}, \vec{v}_4 = \begin{pmatrix}1\\1\\0\end{pmatrix} \in \mathbb{R}^3.$$

These vectors are linearly dependent. Note that $\vec{v}_4 = \vec{v}_1 + \vec{v}_2$. In other words, $\vec{v}_4 \in \operatorname{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$, so by the geometric definition of linear independnece, the vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ are linearly dependent vectors in $\mathbb{R}^3$. However, $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are linearly independent vectors in $\mathbb{R}^3$. We can see this by considering the vector equation $a_1\vec{v}_1 + a_2\vec{v}_2 + a_3\vec{v}_3 = \vec{0}$. This has a corresponding system of linear equations with augmented matrix

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

The matrix is already in RREF and has a pivot in every column except the last, so by Rouché-Capelli (Theorem 1.26), the vector equation

$$a_1\vec{v}_1 + a_2\vec{v}_2 + a_3\vec{v}_3 = \vec{0}$$

has a unique solution, and the vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are linearly independent.

P2.6 (a) This set is linearly dependent, since we have the nontrivial solution $(1, 1, 1)$ to the equation

$$x(\vec{u} - \vec{v}) + y(\vec{v} - \vec{w}) + z(\vec{w} - \vec{u}) = \vec{0}.$$

(b) This set is linearly independent. To see this, consider the vector equation

$$x(\vec{u} + \vec{v}) + y(\vec{v} + \vec{w}) + z(\vec{w} + \vec{u}) = \vec{0}. \tag{C.1}$$

Rearranging gives

$$(x+z)\vec{u} + (x+y)\vec{v} + (y+z)\vec{w} = \vec{0}.$$

But since $\vec{u}, \vec{v}, \vec{w}$ are assumed to be linearly independent, we must have

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

This system has augmented matrix

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

which has reduced row echelon form

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

Hence, $x = y = z = 0$ is the only solution to equation (C.1) and so the vectors $\{\vec{u} + \vec{v}, \vec{v} + \vec{w}, \vec{w} + \vec{u}\}$ are linearly independent.

C.3. Chapter 3 Exercise Solutions

P3.1 Let $\mathcal{S} = \{\vec{v}_1, \ldots, \vec{v}_m\} \subseteq \mathbb{R}^n$ such that $\mathcal{S}$ is linearly independent. We want to show that $m \leq n$. We can prove this by contradiction. Assume, for the sake of contradiction, that $m > n$. Consider the vector equation

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a + m\vec{v}_m = \vec{0}.$$

Looking at the augmented matrix for this system, we see that it has n rows and $m + 1$ columns. Since the last column is all zeros, it cannot be a pivot column, and so the system of equations is consistent. Since there are only n rows, there can be at most n pivot columns. And since the number of columns of the coefficient matrix m is greater than the number of rows, it follows that there is at least one column that will not have a pivot in the RREF of the coefficient matrix. By Rouché-Capelli (Theorem 1.26), the vector equation

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a + m\vec{v}_m = \vec{0}$$

has infinitely many solutions. Then, by the algebraic definition of linear dependence, it follows that the vectors $\vec{v}_1, \ldots, \vec{v}_m$ are linearly dependent. But this contradicts that $\mathcal{S}$ is a set of linearly independent vectors. So our assumption that $m > n$ must be incorrect, and we must have $m \leq n$.

P3.2 Note that there are two directions to prove here. Let's do the easier one first. Let V be a vector space. If $\vec{0} \in V$, then V is most definitely non-empty. Now for the other direction, assume that V is non-empty. Then there exists some element $\vec{v} \in V$. Since V is a vector space, it is closed under scalar multiplication. For any scalar $c \in \mathbb{R}$, we have $c\vec{v} \in V$. In particular, for $c = 0$, we have $0 \cdot \vec{v} = \vec{0} \in V$, which is what we wanted to show.

P3.3 Let $\mathcal{A} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a set of linearly independent vectors in $\mathbb{R}^n$. To show that $\mathcal{A}$ is a basis for $\mathbb{R}^n$, we need to show that $\mathcal{A}$ is a set of linearly independent vectors that span $\mathbb{R}^n$. Since we assume that the vectors in $\mathcal{A}$ are linearly independent, we only need to prove that $\mathcal{A}$ spans $\mathbb{R}^n$. For the sake of contradiction, assume we have a vector $\vec{w} \in \mathbb{R}^n$ that is not in the span of the vectors in $\mathcal{A}$. Then the set $\mathcal{A}' = \{\vec{v}_1, \dots \vec{v}_n, \vec{w}\}$ is a set of $n+1$ linearly independent vectors. By Lemma 3.3, we get that $n+1 \leq n$, which is a contradiction. So $\vec{w}$ must be in $\text{Span}(\vec{v}_1, \dots, \vec{v}_n)$. In other words, $\mathcal{A}$ is a linearly independent generating set for $\mathbb{R}^n$. Thus, $\mathcal{A}$ is a basis for $\mathbb{R}^n$.

P3.4 (a) The set W is a subspace of $\mathbb{R}^3$. First, we show that W is nonempty. Note that for $x = y = 0$, we get $\vec{0} \in W$. Now let $\vec{v}_1, \vec{v}_2 \in W$, then we must have

$$\vec{v}_1 = \begin{pmatrix} 2x_1 + y_1 \\ x_1 - y_1 \\ x_1 + y_1 \end{pmatrix}, \begin{pmatrix} 2x_2 + y_2 \\ x_2 - y_2 \\ x_2 + y_2 \end{pmatrix}$$

for some $x_1, x_2, y_1, y_2 \in \mathbb{R}$. In that case, we have

$$\begin{aligned}
\vec{v}_1 + \vec{v}_2 &= \begin{pmatrix} 2x_1 + y_1 \\ x_1 - y_1 \\ x_1 + y_1 \end{pmatrix} + \begin{pmatrix} 2x_2 + y_2 \\ x_2 - y_2 \\ x_2 + y_2 \end{pmatrix} \\
&= \begin{pmatrix} 2x_1 + y_1 + 2x_2 + y_2 \\ x_1 - y_1 + x_2 - y_2 \\ x_1 + y_1 + x_2 + y_2 \end{pmatrix} \\
&= \begin{pmatrix} 2(x_1 + x_2) + (y_1 + y_2) \\ (x_1 + x_2) - (y_1 + y_2) \\ (x_1 + x_2) + (y_1 + y_2) \end{pmatrix} \\
&= \begin{pmatrix} 2x + y \\ x - y \\ x + y \end{pmatrix}, \text{ where } x = x_1 + x_2, y = y_1 + y_2 \in \mathbb{R}.
\end{aligned}$$

It follows that $\vec{v}_1 + \vec{v}_2 \in W$ and W is closed under vector addition. Now let $c \in \mathbb{R}$ be a scalar. We have

$$\begin{aligned}
c\vec{v}_1 &= c \begin{pmatrix} 2x_1 + y_1 \\ x_1 - y_1 \\ x_1 + y_1 \end{pmatrix} \\
&= \begin{pmatrix} 2cx_1 + cy_1 \\ cx_1 - cy_1 \\ cx_1 + cy_1 \end{pmatrix} \\
&= \begin{pmatrix} 2x + y \\ x - y \\ x + y \end{pmatrix}, \text{ where } x = cx_1, y = cy_1 \in \mathbb{R}.
\end{aligned}$$

Therefore, $c\vec{v}_1 \in W$, and W is closed under scalar multiplication. Since W is a nonempty subset of $\mathbb{R}^3$ that is closed under addition and scalar multiplication, it follows that W is a subspace of $\mathbb{R}^n$.

(b) The set W is not a subspace of $\mathbb{R}^2$. Note that the vector $\begin{pmatrix} 1 \\ -1 \end{pmatrix} \in W$, but its scalar multiple $0 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \notin W$. Since W is not closed under scalar multiplication, it is not a subspace of $\mathbb{R}^2$.

(c) The set W is not a subspace of $\mathbb{R}^3$. Consider the vectors $\begin{pmatrix}1\\0\\0\end{pmatrix}$, which you get when $x = y = 1, z = 0$, and $\begin{pmatrix}0\\1\\0\end{pmatrix}$, which you get when $x = z = 1, y = 0$. If we add these two vectors, we get $\begin{pmatrix}1\\0\\0\end{pmatrix} + \begin{pmatrix}0\\1\\0\end{pmatrix} = \begin{pmatrix}1\\1\\0\end{pmatrix}$. The resulting vector cannot be in W. If we have $xy = 1, xz = 1, yz = 0$, then $yz = 0$ implies that $y = 0$ or $z = 0$ or both. But this would imply that one of xy, xz would also be zero, which is not the case. Since W is not closed under vector addition, it follows that W is not a subspace of $\mathbb{R}^3$.

P3.5 Since V and W are both vector subspaces of $\mathbb{R}^n$, they are both nonempty and there are some vectors $\vec{v} \in V$, and $\vec{w} \in W$. This means that $\vec{v} + \vec{w} \in V + W$ proving that V+W is non-empty. Now suppose $\vec{u}_1, \vec{u}_2 \in V + W$. Then $\vec{u}_1 = \vec{v}_1 + \vec{w}_1$ and $\vec{u}_2 = \vec{v}_2 + \vec{w}_2$ for some vectors $\vec{v}_1, \vec{v}_2 \in V, \vec{w}_1, \vec{w}_2 \in W$. Now

$$\begin{aligned}\vec{u}_1 + \vec{u}_2 &= \vec{v}_1 + \vec{w}_1 + \vec{v}_2 + \vec{w}_2\\ &= (\vec{v}_1 + \vec{v}_2) + (\vec{w}_1 + \vec{w}_2)\\ &= \vec{v} + \vec{w}, \text{where } \vec{v} = \vec{v}_1 + \vec{v}_2, \vec{w} = \vec{w}_1 + \vec{w}_2.\end{aligned}$$

Since V and W are subspaces of $\mathbb{R}^n$, they are closed under vector addition, so $\vec{v} \in V$ and $\vec{w} \in W$. It follows that $\vec{u}_1 + \vec{u}_2 \in V + W$ and $V + W$ is closed under vector addition.

Now let c be a scalar, then

$$\begin{aligned}c\vec{u}_1 &= c(\vec{v}_1 + \vec{w}_1)\\ &= c\vec{v}_1 + c\vec{w}_1\\ &= \vec{v} + \vec{w}, \text{where } \vec{v} = c\vec{v}_1, \vec{w} = c\vec{w}_1.\end{aligned}$$

Since V and W are subspaces of $\mathbb{R}^n$, they are closed under vector scalar multiplication, so $\vec{v} \in V$ and $\vec{w} \in W$. It follows that $c\vec{u}_1 \in V + W$ and $V + W$ is closed under scalar multiplication. Since $V + W$ is a nonempty subset of $\mathbb{R}^n$ that is closed under vector addition and scalar multiplication, $V + W$ is a vector subspace of $\mathbb{R}^n$

P3.6 The statement is true. Let W and V be subspaces of $\mathbb{R}^n$. Then W and V are nonempty subspaces of $\mathbb{R}^n$. Let $\vec{v} \in V$, and $\vec{w} \in W$. Since V and W are closed under scalar multiplication, $0\vec{v} = \vec{0} \in V$ and $0\vec{w} = \vec{0} \in W$, and so $\vec{0} \in W \cap V$, and $W \cap V$ is nonempty. Now suppose $\vec{x}_1, \vec{x}_2 \in W \cap V$ and let c be a scalar. Then $\vec{x}_1, \vec{x}_2 \in W$ and $\vec{x}_1, \vec{x}_2 \in V$. Since W and V are subspaces of $\mathbb{R}^n$, they are closed under addition, and $\vec{x}_1 + \vec{x}_2 \in W$ and $\vec{x}_1 + \vec{x}_2 \in V$. So $\vec{x}_1 + \vec{x}_2 \in W \cap V$, and $W \cap V$ is closed under vector addition. Additionally, since W and V are subspaces of $\mathbb{R}^n$, they are closed under scalar multiplication, and so $c\vec{x}_1 \in W$ and $c\vec{x}_1 \in V$. Therefore, $c\vec{x}_1 \in W \cap V$, and $W \cap V$ is closed under scalar multiplication. Since $W \cap V$ is a nonempty subspace of $\mathbb{R}^n$ that is closed under vector addition and scalar multiplication, it follows that $W \cap V$ is a subspace of $\mathbb{R}^n$.

P3.7 The statement is false. Let $W = \text{Span}\left(\begin{pmatrix}1\\0\end{pmatrix}\right)$, $V = \text{Span}\left(\begin{pmatrix}0\\1\end{pmatrix}\right)$. Note that W and V are subspaces of $\mathbb{R}^2$ by Theorem 3.4. Now consider $W \cup V$. Take the vectors $\vec{w} = \begin{pmatrix}1\\0\end{pmatrix} \in W$, and $\vec{v} = \begin{pmatrix}0\\1\end{pmatrix} \in V$, and note that $\vec{w}, \vec{v} \in W \cup V$. However, $\vec{w} + \vec{v} = \begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix} \notin W$, and $\vec{w} + \vec{v} = \begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix} \notin V$, so that $\vec{w} + \vec{v} \notin W \cup V$. Since $W \cup V$ is not closed under vector addition, it is not a subspace of $\mathbb{R}^2$.

P3.8 Let $\vec{v}$ in $\mathbb{R}^n$ such that $\vec{v} \notin V$, and let $B = \{\vec{v}_1, \dots \vec{v}_m\}$. For the sake of contradiction, assume that the set $B \cup \{\vec{v}\}$ is linearly dependent. This means that there is a nontrivial solution to the equation

$$a_1\vec{v}_1 + \cdots + a_m\vec{v}_m + a_{m+1}\vec{v} = \vec{0}.$$

We have two cases, either $a_{m+1} = 0$ or $a_{m+1} \neq 0$. If $a_{m+1} = 0$, then there is a nontrivial solution to the vector equation

$$a_1\vec{v}_1 + \cdots + a_m\vec{v}_m = \vec{0}.$$

But this implies that the vectors in B are linearly independent contradicting the assumption that B is a basis for V. So we must have $a_{m+1} \neq 0$. In that case, by rearranging the vector equation, we have

$$\tfrac{a_1}{a_{m+1}}\vec{v}_1 + \cdots + \tfrac{a_m}{a_{m+1}}\vec{v}_m = \vec{v}.$$

But this implies that $\vec{v} \in \text{Span}(\vec{v_1}, \dots \vec{v}_m) = V$, which contradicts the assumption that $\vec{v} \notin V$. It follows that $B \cup \{\vec{v}\}$ must be linearly independent.

P3.9 Let V be a vector subspace of $\mathbb{R}^n$ of dimension $m \geq 1$, and let $B = \{\vec{b}_1, \vec{b}_2, \dots, \vec{b}_m\}$ be a basis for V. If $m = n$, we are done, since then $V = \mathbb{R}^n$. Assume $m < n$, then there exists some vector $v_{m+1} \in \mathbb{R}^n$ such that $\vec{v}_{m+1} \notin V$. By the exercise above, $B \cup \{v_{m+1}\}$ is linearly independent. If $n = m+1$, we're done, because by a previous exercise, any set of n linearly independent vectors in $\mathbb{R}^n$ forms a basis for $\mathbb{R}^n$. Otherwise, we keep repeating this process (finding vectors $\vec{v}_{m+i} \notin \text{Span}(\vec{b}_1, \dots, \vec{b}_m, \vec{v}_{m+1}, \dots, \vec{v}_{m+i-1})$) until we find a set C of n linearly independent vectors. Then C will be a basis for $\mathbb{R}^n$ and we will have $B \subseteq C$.

P3.10 We claim that $\{\vec{v}_1, \vec{v}_2, \vec{w}_1, \vec{w}_2\}$ is a basis for $V + W$. First, we show that the vectors are linearly independent. Consider the vector equation

$$a_1\vec{v}_1 + a_2\vec{v}_2 + b_1\vec{w}_1 + b_2\vec{w}_2 = \vec{0}.$$

Rearranging, we get

$$a_1\vec{v}_1 + a_2\vec{v}_2 = -b_1\vec{w}_1 - b_2\vec{w}_2.$$

Now $a_1\vec{v}_1 + a_2\vec{v}_2 \in V$, and $-b_1\vec{w}_1 - b_2\vec{w}_2 \in W$, so this would mean that $a_1\vec{v}_1 + a_2\vec{v}_2, -b_1\vec{w}_1 - b_2\vec{w}_2 \in V \cap W = \{\vec{0}\}$. And so we have

$$a_1\vec{v}_1 + a_2\vec{v}_2 = \vec{0},$$
$$-b_1\vec{w}_1 - b_2\vec{w}_2 = \vec{0}.$$

Since $\{\vec{v}_1, \vec{v}_2\}$ is a basis for V and $\{\vec{w}_1, \vec{w}_2\}$ is a basis for W, these two sets of vectors are linearly independent, and so the last two vector equations only have the trivial solution. In other words, we must have $a_1 = a_2 = b_1 = b_2 = 0$.

This means that the original vector equation above only has the trivial solution and the vectors $\vec{v}_1, \vec{v}_2, \vec{w}_1, \vec{w}_2$ are linearly independent.

Next, we show that $\{\vec{v}_1, \vec{v}_2, \vec{w}_1, \vec{w}_2\}$ is a generating set for $V + W$. Let $\vec{x} \in V + W$, then $\vec{x} = \vec{v} + \vec{w}$ for some $\vec{v} \in V, \vec{w} \in W$. Since $\{\vec{v}_1, \vec{v}_2\}$ is a basis for V, there are scalars c_1, c_2 such that $\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2$. Since $\{\vec{w}_1, \vec{w}_2\}$ is a basis for W, there are scalars d_1, d_2 such that $\vec{w} = d_1\vec{w}_1 + d_2\vec{w}_2$. Then $\vec{x} = \vec{v} + \vec{w} = c_1\vec{v}_1 + c_2\vec{v}_2 + d_1\vec{w}_1 + d_2\vec{w}_2 \in \operatorname{Span}(\vec{v}_1, \vec{v}_2, \vec{w}_1, \vec{w}_2)$. This shows that $\{\vec{v}_1, \vec{v}_2, \vec{w}_1, \vec{w}_2\}$ is a generating set for $V + W$, and we've shown that the vectors are linearly independent, and so $\{\vec{v}_1, \vec{v}_2, \vec{w}_1, \vec{w}_2\}$ is a basis for $V + W$.

P3.11 The statement is false. Consider $V = W = \operatorname{Span}\left(\begin{pmatrix}1\\0\end{pmatrix}\right)$. By Theorem 3.4, we know that V and W are subspaces of $\mathbb{R}^2$. Since $\left\{\begin{pmatrix}1\\0\end{pmatrix}\right\}$ is a linearly independent generating set for V and W, it is also a basis for V and W. This means that $\dim(V) = \dim(W) = 1$. Now

$$\begin{aligned}
V + W &= \{\vec{v} + \vec{w} : \vec{v} \in V, \vec{w} \in W\} \\
&= \{\vec{v} + \vec{w} : \vec{v}, \vec{w} \in V\} && \text{since } V = W \\
&= \{\vec{v} : \vec{v} \in V\} && \text{since } V \text{is a subspace, closed under addition} \\
&= V.
\end{aligned}$$

And so, $\dim(V + W) = \dim(V) = 1$. However $\dim(V) + \dim(W) = 1 + 1 = 2 \neq \dim(V + W)$.

P3.12 It is not possible. For the sake of contradiction, suppose that there are two vector subspaces V and W of $\mathbb{R}^3$ with $V \cap W = \{\vec{0}\}$ so that $\dim V = \dim W = 2$. Let $\{\vec{v}_1, \vec{v}_2\}$ be a basis for V and let $\{\vec{w}_1, \vec{w}_2\}$ be a basis for W. Note that $V + W$ is a subspace of $\mathbb{R}^3$ as shown in a previous exercise, and that $\{\vec{v}_1, \vec{v}_2, \vec{w}_1, \vec{w}_2\}$ is a basis for $V + W$ as we found in the solution of a previous exercise. So we know that $\dim(V + W) = 4$, but that's not possible because $V + W$ is a subspace of $\mathbb{R}^3$ and $\dim(\mathbb{R}^3) = 3 < 4$. It follows that there cannot be two vector subspaces V and W of $\mathbb{R}^3$ with $V \cap W = \{\vec{0}\}$ so that $\dim V = \dim W = 2$.

C.4. Chapter 4 Exercise Solutions

P4.1 It must be the case that $c = -1$. If the column space of the matrix has basis $\left\{\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}2\\1\\2\end{pmatrix}\right\}$, then column 1 and column 2 are linearly independent and their span is equal to the span of all three column vectors. This means that column 3 is a linear combination of column 1 and column 2. In particular, $\begin{pmatrix}-1\\1\\c\end{pmatrix} = a\begin{pmatrix}1\\1\\1\end{pmatrix} + b\begin{pmatrix}2\\1\\2\end{pmatrix}$. Expanding this, we get $a + 2b = -1, a + b = 1$, and $a + 2b = c$. From the first and third equations, we can see that we must have $c = -1$.

P4.2 Recall that the null space of a matrix A are the solutions to the vector equation $A\vec{x} = \vec{0}$. If the null space of the matrix has basis $\left\{ \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \right\}$, then in particular,

$$\begin{pmatrix} 2 & 1 & a \\ 1 & 2 & b \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Computing the product of the matrix and vector, and equating them to the corresponding entries in the zero vector, we get

$$4 - 1 + a = 0,$$

$$2 - 2 + b = 0.$$

And so we must have $a = -3$, and $b = 0$.

P4.3 We know that our matrix, A must have exactly four columns in order to have a null space that is a subspace of $\mathbb{R}^4$. Since there's only one vector in the basis for $\mathrm{Nul}(A)$, we know that $\mathrm{nullity}(A) = 1$. By Theorem 4.7, we know that A has one non-pivot column, and $4 - 1 = 3$ pivot columns. This means that our matrix must have at least 3 rows (to accommodate 3 pivot columns). Let $\vec{r} = \begin{pmatrix} a & b & c & d \end{pmatrix}$ be a row of our matrix. Since $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix} \right\}$ is a basis for the null space of A, then

$$A \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \vec{0}.$$

The entry of $A \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix}$ corresponding to row $\vec{r}$ is $a + b + c - d$. From the equation above, we have $a + b + c - d = 0$. One of the simplest ways to build up our matrix A is to have A in RREF, with 3 rows, and select our rows one-by-one, making sure each row has one leading one, and entries that satisfy the equation $a + b + c - d = 0$. The following is one of many options for A,

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

P4.4 The statement is false. For the sake of contradiction, let A be a 3×3 matrix, whose column space and null space have the same dimension k. Then we have $\dim(\mathrm{Col}(A)) = \mathrm{rank}(A) = k$ and $\dim(\mathrm{Nul}(A)) = \mathrm{nullity}(A) = k$. According to the Rank-Nullity Theorem (Corollary 4.8), $\mathrm{rank}(A) + \mathrm{nullity}(A) = 3$. But this implies that $2k = 3$, or that $k = \frac{3}{2}$, which contradicts the fact that the rank and nullity of a matrix are non-negative integers (since they are the number of pivot, or non-pivot columns of a matrix). Thus, a 3×3 matrix, whose column space and null space have the same dimension does not exist.

P4.5 The statement is true. Let A be an $n \times n$ matrix of rank r. By Theorem 4.7, the nullity of A is the number of non-pivot columns of A, $n - r$, which is the number of non-pivot columns in $\text{rref}(A)$. Since $\text{rank}(A) = r$, the $\text{rref}(A)$ has r pivots in its first r rows. The remaining $n - r$ rows must be all zeros (otherwise, they'd contain a pivot). And so the nullity of A is the number of all zero rows in $\text{rref}(A)$.

P4.6 The statement is false. Let $A = B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Note that A and B are in reduced row echelon form and they have a pivot in each column, and so $\text{rank}(A) = \text{rank}(B) = 2$. Now $A + B = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, and $\text{rref}(A + B) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Since $\text{rref}(A + B)$ has 2 pivot columns, then $\text{rank}(A + B) = 2 \neq 4 = 2 + 2 = \text{rank}(A) + \text{rank}(B)$.

P4.7 Consider the system of equations $(A-B)\vec{x} = \vec{0}$. Since $(A-B)\vec{x} = A\vec{x}-B\vec{x} \neq \vec{0}$ for any non-zero $\vec{x} \in \mathbb{R}^n$, we know that the system $(A - B)\vec{x} = \vec{0}$ has only the trivial solution. This means that $\text{Nul}(A - B) = \{\vec{0}\}$ and $\text{nullity}(A - B) = 0$. By the Rank-Nullity Theorem,

$$\begin{aligned} \text{rank}(A - B) &= n - \text{nullity}(A - B) \\ &= n - 0 \\ &= n. \end{aligned}$$

C.5. Chapter 5 Exercise Solutions

P5.1 Let F be a linear transformation. Then for any scalar c, and any vector $\vec{v}$ in the domain of F, we have $F(c\vec{v}) = cF(\vec{v})$. In particular, this is true for $c = 0$, and we get $F(\vec{0}) = F(0\vec{v}) = 0F(\vec{v}) = \vec{0}$, as required.

P5.2 First, we show that $F(V)$ is nonempty. Since V is a vector subspace of $\mathbb{R}^n$, V is nonempty. So there must be some vector $\vec{v} \in V$, and its image $F(\vec{v}) \in F(V)$. Next, we show that $F(V)$ is closed under vector addition and scalar multiplication. Let $\vec{x_1}, \vec{x}_2 \in F(V)$, and let c be a scalar. Then $\vec{x}_1 = F(\vec{v}_1), \vec{x}_2 = F(\vec{v}_2)$, for some vectors $\vec{v}_1, \vec{v}_2 \in V$. Since V is a vector subspace, it is closed under vector addition, and $\vec{v}_1 + \vec{v}_2 \in V$. Therefore, we have $F(\vec{v}_1 + \vec{v}_2) \in F(V)$. But F is a linear transformation, so

$$F(\vec{v}_1 + \vec{v}_2) = F(\vec{v}_1) + F(\vec{v}_2) = \vec{x}_1 + \vec{x}_2 \in F(V).$$

It follows that $F(V)$ is closed under vector addition. Additionally, since V is a vector subspace, it is closed under scalar multiplication, and $c\vec{v}_1 \in V$. Therefore, we have $F(c\vec{v}_1) \in F(V)$. But F is a linear transformation, so

$$F(c\vec{v}_1) = cF(\vec{v}_1) = c\vec{x}_1 \in F(V).$$

It follows that $F(V)$ is closed under scalar multiplication. We conclude that $F(V)$ is a vector subspace of $\mathbb{R}^m$ by definition.

P5.3 Let $F : \mathbb{R}^n \to \mathbb{R}^k$ and $G : \mathbb{R}^k \to \mathbb{R}^m$ be linear functions. Let $\vec{u}, \vec{v} \in \mathbb{R}^n$, and let c be a scalar. Now consider the following,

$$\begin{aligned} G \circ F(\vec{u} + \vec{v}) &= G(F(\vec{u} + \vec{v})) \\ &= G(F(\vec{u}) + F(\vec{v})) && \text{since } F \text{ is linear} \\ &= G(F(\vec{u})) + G(F(\vec{v})) && \text{since } G \text{ is linear} \\ &= G \circ F(\vec{u}) + G \circ F(\vec{v}). \end{aligned}$$

We also have,

$$\begin{aligned} G \circ F(c\vec{u}) &= G(F(c\vec{u})) \\ &= G(cF(\vec{u})) && \text{since } F \text{ is linear} \\ &= cG(F(\vec{u})) && \text{since } G \text{ is linear} \\ &= cG \circ F(\vec{u}). \end{aligned}$$

It follows that $G \circ F$ is also linear.

P5.4 The statement is False. Let $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, then clearly A is not equal to the zero matrix. However $A^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ is equal to the zero matrix.

P5.5 The statement is False. Let $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, then $A^2 = 0$. And let $B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, then $B^2 = 0$. By definition, A and B are both nilpotent matrices. Now consider their product, $AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. We claim that AB is not nilpotent. In fact, we have $(AB)^2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = AB$. This means that raising AB to any power will result in AB again, in other words, $(AB)^n = AB \neq 0$ for any positive integer n. Thus AB is not nilpotent.

P5.6 Let A and B be $n \times n$ matrices, and let $B = \begin{pmatrix} \vec{b}_1 & \vec{b}_2 & \dots & \vec{b}_n \end{pmatrix}$. Note that $AB = \begin{pmatrix} A\vec{b}_1 & A\vec{b}_2 & \dots & A\vec{b}_n \end{pmatrix}$. This means that every column of AB is a linear combination of the columns of A. It follows that $\operatorname{rank}(AB) \leq \operatorname{rank}(A)$. Next, we show that $\operatorname{rank}(AB) \leq \operatorname{rank}(B)$. We can view AB as a composition of linear maps. Note that $\ker(B) \subseteq \ker(AB)$, this is because if $\vec{x} \in \ker(B)$, then $AB\vec{x} = A(B\vec{x}) = A\vec{0} = \vec{0}$, and $\vec{x} \in \ker(AB)$. Therefore, $\operatorname{nullity}(B) \leq \operatorname{nullity}(AB)$. Using the Rank-Nullity Theorem, we know that $\operatorname{rank}(B) + \operatorname{nullity}(B) = n = \operatorname{rank}(AB) + \operatorname{nullity}(AB)$. And we get that $\operatorname{rank}(AB) \leq \operatorname{rank}(B)$. Combining our two results, we've shown that $\operatorname{rank}(AB) \leq \min\{\operatorname{rank}(A), \operatorname{rank}(B)\}$.

P5.7 The statement is True. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a one-to-one linear transformation. First, we claim that the nullity of T must be 0. From a previous exercise, we know that $T(\vec{0}) = \vec{0}$, so $\vec{0} \in \ker(T)$. We want to show that there is no other vector in the kernel of T. Suppose $\vec{x} \neq \vec{0}$ is in the kernel of T. This means that $T(\vec{x}) = \vec{0} = T(\vec{0})$, but this contradicts that T is one-to-one. Therefore, $\ker(T) = \{\vec{0}\}$, and $\operatorname{nullity}(T) = 0$. By the Rank-Nullity Theorem, we know that $\text{rank}(T) = n - \text{nullity}(T) = n - 0 = n$. But $\text{rank}(T) = \dim(\text{im}(T))$, so the image of T, $T(\mathbb{R}^n) \subseteq \mathbb{R}^n$, has dimension n. This can only happen if $T(\mathbb{R}^n) = \mathbb{R}^n$, and so T must also be onto.

P5.8 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Suppose $n < m$. We know that F has a defining matrix, A_F of size $m \times n$, and $\text{im}(F) = \text{Col}(A_F)$ by Activity 5.8. Now $\text{rank}(A_F) \leq \min(m, n)$ (by Ex1.1), and since $n < m$, we have $\text{rank}(A_F) \leq n < m$. So the dimension of the image of F is strictly less than the dimension of its codomain ($\mathbb{R}^m$). It follows that F is not surjective.

P5.9 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Suppose $n > m$. We know that F has a defining matrix, A_F of size $m \times n$, and $\ker(F) = \text{Nul}(A_F)$ by Activity 5.8. Now $\text{rank}(A_F) \leq \min(m, n)$ (by Ex1.1), and since $n > m$, we have $\text{rank}(A_F) \leq m < n$. By the Rank-Nullity Theorem, $\text{rank}(A_F) + \text{nullity}(A_f) = n$. Since $\text{rank}(A_F) < n$, it follows that $\text{nullity}(A_F) > 0$. This means that there is a non-zero vector $\vec{x} \in \text{Nul}(A_F) = \ker(T)$, and $T(\vec{x}) = \vec{0}$. But we know from the first exercise in this section, that $T(\vec{0}) = \vec{0}$. It follows that T is not injective.

P5.10 Let $F : V \to W$ be a linear transformation between vector spaces V and W. Assume F is injective. By a previous exercise, we know that $F(\vec{0}) = \vec{0}$, and since F is injective, no other vector can map to $\vec{0}$. So $\ker F = \{\vec{0}\}$. Now to prove the converse, suppose that $\ker(F) = \{\vec{0}\}$, and that $F(\vec{x}) = F(\vec{y})$ for some vectors $\vec{x}, \vec{y} \in V$. Then

$$\begin{aligned} \vec{0} &= F(\vec{x}) - F(\vec{y}) \\ &= F(\vec{x} - \vec{y}) && \text{by the linearity of } F. \end{aligned}$$

It follows that $\vec{x} - \vec{y} \in \ker(F) = \{\vec{0}\}$. Therefore, $\vec{x} - \vec{y} = \vec{0}$, or $\vec{x} = \vec{y}$, which shows that F is injective.

P5.11 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and $\{\vec{v}_1, \ldots, \vec{v}_n\}$ a basis for $\mathbb{R}^n$. Suppose F is injective, and consider the vector equation,

$$a_1 F(\vec{v}_1) + \cdots + a_n F(\vec{v}_n) = \vec{0}.$$

By the linearity of F, this can be rewritten as

$$F(a_1 \vec{v}_1 + \cdots + a_n \vec{v}_n) = \vec{0}.$$

Since F is injective and $F(\vec{0}) = \vec{0}$ (by a previous Exercise), it follows that we must have $a_1 \vec{v}_1 + \cdots + a_n \vec{v}_n = \vec{0}$. Now $\{\vec{v}_1, \ldots, \vec{v}_n\}$ a basis for $\mathbb{R}^n$ and so these vectors must be linearly independent and therefore, the vector equation $a_1 \vec{v}_1 + \cdots + a_n \vec{v}_n = \vec{0}$ only has the trivial solution, in other words, we must have $a_1 = \cdots = a_n = 0$. It follows that the vectors $F(\vec{v}_1), \ldots, F(\vec{v}_n)$ are linearly independent.

Next, we prove the converse. Suppose that the set $\{F(\vec{v}_1), \ldots, F(\vec{v}_n)\}$ is linearly independent. Let $\vec{x} = a_1 \vec{v}_1 + \cdots + a_n \vec{v}_n \in \ker F \subseteq \mathbb{R}^n$. Then

$$\begin{aligned} \vec{0} &= F(\vec{x}) && \text{since } \vec{x} \in \ker F \\ &= F(a_1 \vec{v}_1 + \cdots + a_n \vec{v}_n) \\ &= a_1 F(\vec{v}_1) + \cdots + a_n F(\vec{v}_n) && \text{by the linearity of } F. \end{aligned}$$

Since the vectors $F(\vec{v}_1), \ldots, F(\vec{v}_n)$ are linearly independent, it follows that the vector equation above only has the trivial solution. In other words, $a_1 = \cdots = a_n = 0$, and so $\vec{x} = 0\vec{v}_1 + \cdots + 0\vec{v}_n = \vec{0}$. But then $\ker(F) = \{\vec{0}\}$, and by the previous exercise, F is injective.

P5.12 Let A_F be the defining matrix for F. Since F is surjective, we know that $\operatorname{im}(F) = \mathbb{R}^3$. But $\operatorname{im}(F) = \operatorname{Col}(A_F)$ (by Activity 5.8), so $\operatorname{rank}(A_F) = \dim(\operatorname{Col}(A_F) = 3$. By the Rank-Nullity Theorem, $\operatorname{nullity}(A_F) = 7 - \operatorname{rank}(A_F) = 7 - 3 = 4$. But $\ker(F) = \operatorname{Nul}(A_F)$, and so the dimension of the kernel of F is equal to the dimension of the null space of A_F, which is just the nullity of A_F. Therefore, the dimension of $\ker(F)$ is 4.

P5.13 Let $F : \mathbb{R}^5 \to \mathbb{R}^4$ be a linear transformation with $\operatorname{nullity}(F) = 2$. Then F has a 4×5 defining matrix A_F with $\operatorname{nullity}(A_F) = 2$. By the Rank-Nullity Theorem, $\operatorname{rank}(A_F) + \operatorname{nullity}(A_F) = 5$, which implies that $\operatorname{rank}(A_F) = 5 - 2 = 3$. But the rank of A_F is equal to the dimension of the column space of A_F, which is equal to the image of F by activity 5.8. Hence, $\dim(\operatorname{im}(F)) = 3$. Now Theorem 6.6 states that any two vector spaces with the same dimension are isomorphic. Since $\dim(\operatorname{im}(F)) = 3 = \dim(\mathbb{R}^3)$, it follows that $\operatorname{im}(F)$ is isomorphic to $\mathbb{R}^3$.

P5.14 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be an injective linear transformation with $n \leq m$, and suppose that V is a vector subspace of $\mathbb{R}^n$ with $\dim(V) = k$. Let $\{\vec{v}_1, \ldots, \vec{v}_k\}$ be a basis for V. We claim that $\{T(\vec{v}_1), \ldots, T(\vec{v}_k)\}$ is a basis for $T(V)$. First, let $\vec{x} \in T(V)$, then $\vec{x} = T(\vec{u})$ for some $\vec{u} \in V$. Now $\vec{u} = a_1\vec{v}_1 + \cdots + a_k\vec{v}_k$ for some scalars $a_1, \ldots, a_n$ (because $\{\vec{v}_1, \ldots, \vec{v}_k\}$ be a basis for V). Also,

$$
\begin{aligned}
\vec{x} &= T(\vec{u}) \\
&= T(a_1\vec{v}_1 + \cdots + a_k\vec{v}_k) \\
&= a_1T(\vec{v}_1) + \cdots + a_kT(\vec{v}_k) && \text{by the linearity of } T.
\end{aligned}
$$

This shows that $\{T(\vec{v}_1), \ldots, T(\vec{v}_k)\}$ is a spanning set for $T(V)$. Next, we show that the set is linearly independent. Consider the vector equation,

$$a_1T(\vec{v}_1) + \cdots + a_kT(\vec{v}_k) = \vec{0}.$$

By the linearity of T, this can be rewritten as

$$T(a_1\vec{v}_1 + \cdots + a_k\vec{v}_k) = \vec{0}.$$

Since T is injective and $T(\vec{0}) = \vec{0}$ (by a previous Exercise), it follows that we must have $a_1\vec{v}_1 + \cdots + a_k\vec{v}_k = \vec{0}$. Now $\{\vec{v}_1, \ldots, \vec{v}_k\}$ a basis for V and so these vectors must be linearly independent and therefore, the vector equation $a_1\vec{v}_1 + \cdots + a_k\vec{v}_k = \vec{0}$ only has the trivial solution, in other words, we must have $a_1 = \cdots = a_k = 0$. It follows that the vectors $T(\vec{v}_1), \ldots, T(\vec{v}_k)$ are linearly independent.

Since $\{T(\vec{v}_1), \ldots, T(\vec{v}_k)\}$ is a linearly independent spanning set of $T(V)$, it is, by definition, a basis for $T(V)$. Therefore, $\dim(T(V)) = k$, and by Theorem 6.6, $T(V)$ is isomorphic to $\mathbb{R}^k$.

C.6. Chapter 6 Exercise Solutions

P6.1 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a bijective linear transformation, and let A be the defining matrix for F. Then A is an $m \times n$ matrix. Since F is bijective, it is both injective and surjective. By Theorem 6.3, since F is injective, $\operatorname{rref}(A)$ has a pivot in every column. This implies that A has n pivots, which is the number of columns of A. Also by Theorem 6.3, since F is surjective, $\operatorname{rref}(A)$ has a pivot in every row. It follows that A has m pivots, since it has m rows. But then we must have that $n = m$ as required.

P6.2 It is not possible for an $m \times n$ matrix to have an inverse if $m \neq n$. Let A be an $m \times n$ matrix. Then the transformation T_A defined by A is from $\mathbb{R}^n$ to $\mathbb{R}^m$. Suppose A has an inverse A^{-1}. Then $AA^{-1} = I_m$, where A^{-1} is the defining matrix for $(T_A)^{-1} : \mathbb{R}^m \to \mathbb{R}^n$. It follows that A has size $n \times m$. Now A^{-1} is the inverse of A, so it must also satisfy $A^{-1}A = I_n \neq I_m$. But this does not satisfy the definition of an inverse matrix. Therefore, A cannot have an inverse.

P6.3 Consider $\left(A \mid I_2\right)$, we have the following cases.

Case 1: When $a \neq 0$. Then

$$\left(\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ c & d & 0 & 1 \end{array}\right), \quad \text{dividing } R_1 \text{ by } a$$

$$\sim \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & d - \frac{bc}{a} & -\frac{c}{a} & 1 \end{array}\right), \quad \text{subtracting } cR_1 \text{ from}$$

$$\sim \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right), \quad \text{dividing } R_2 \text{ by } d - \frac{bc}{a} = \frac{ad-bc}{a}, \text{ assuming } ad - bc$$

$$\left(\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & \frac{b}{ad-bc} \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right), \text{ subtracting } \frac{b}{a}R_2 \text{ from } R_1.$$

Note that if $ad - bc = 0$, then we would have no pivot in the second column, and so $\operatorname{rref}(A) \neq I_2$, and by Theorem 6.21, A would not be invertible. If $ad - bc \neq 0$, then by Theorem 6.17, we get that $A^{-1} = \begin{pmatrix} \frac{d}{ad-bc} & \frac{b}{ad-bc} \\ -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix} = \frac{1}{ad-bc}\begin{pmatrix} d & b \\ -c & a \end{pmatrix}$.

Case 2: When $a = 0$. If $c = 0$, then $ad - bc = 0$. In this case, there can never be a pivot in the first column of $\operatorname{rref}(A)$, so $\operatorname{rref}(A) \neq I_2$, and A will not be invertible by Theorem 6.21. So assume $c \neq 0$. If $b = 0$, then $ad - bc = 0$, and there can never be a pivot in the first row of $\operatorname{rref}(A)$. So we would have $\operatorname{rref}(A) \neq I_2$, and A will not be invertible, again by Theorem 6.21. So assume $b \neq 0$. Then

$$\left(\begin{array}{cc|cc} 0 & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right) \sim \left(\begin{array}{cc|cc} c & d & 0 & 1 \\ 0 & b & 1 & 0 \end{array}\right), \quad \text{swapping } R_1 \text{ and } R_2$$

$$\sim \left(\begin{array}{cc|cc} 1 & \frac{d}{c} & 0 & \frac{1}{c} \\ 0 & b & 1 & 0 \end{array}\right), \quad \text{dividing } R_1 \text{ by } c \neq 0$$

$$\sim \left(\begin{array}{cc|cc} 1 & \frac{d}{c} & 0 & \frac{1}{c} \\ 0 & 1 & \frac{1}{b} & 0 \end{array}\right), \quad \text{dividing } R_2 \text{ by } b \neq 0$$

$$\sim \left(\begin{array}{cc|cc} 1 & 0 & \frac{-d}{bc} & \frac{1}{c} \\ 0 & 1 & \frac{1}{b} & 0 \end{array}\right), \quad \text{subtracting } \frac{d}{c}R_2 \text{ from } R_1.$$

Then by Theorem 6.17 $A^{-1} = \begin{pmatrix} \frac{-d}{bc} & \frac{1}{c} \\ \frac{1}{b} & 0 \end{pmatrix}$. Note that $a = 0$, and so $\frac{-d}{bc} = \frac{d}{ad-bc} = \frac{d}{-bc}$. Also $\frac{1}{c} = \frac{-b}{ad-bc} = \frac{-b}{-bc}$. And $\frac{1}{b} = \frac{-c}{ad-bc} = \frac{-c}{-bc}$. Therefore, $A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ as required.

P6.4 Suppose that B and C are two matrices satisfying

$$AB = BA = I_n$$
$$AC = CA = I_n.$$

This gives

$$B = BI_n = B(AC) = (BA)C = I_nC = C$$

and so $B = C$.

P6.5 Let A be an $n \times n$ matrix.

First we prove (a) $\Leftrightarrow$ (b). By Theorem 6.21, A is invertible if and only if $\text{rref}(A) = I_n$. But $\text{rref}(A) = I_n$ if and only if $\text{rref}(A)$ has n pivots, since A is $n \times n$.

Next we prove (b) $\Leftrightarrow$ (e). If $\text{rref}(A)$ has n pivots, then it has a pivot in every column. By Theorem 3.11, the pivot columns of A, which are all the columns of A, form a basis for $\text{Col}(A)$. But this means that the columns of A are linearly independent. For the converse, assume that the columns of A are linearly independent. Let $A = \begin{pmatrix} \vec{v}_1 & \dots & \vec{v}_n \end{pmatrix}$. Since the $\vec{v}_1, \dots, \vec{v}_n$ are linearly independent, the vector equation

$$c_1\vec{v}_1 + \cdots + c_n\vec{v}_n = \vec{0}$$

has only the trivial solution. But by Rouché-Capelli, this can only happen if $\text{rref}(A)$ has a pivot in every column. Since A is $n \times n$, this would mean that $\text{rref}(A)$ must have n pivots.

Next we prove (b) $\Leftrightarrow$ (g) and (b) $\Leftrightarrow$ (h). This follows directly from Theorem 6.3, because $\text{rref}(A)$ has a pivot in every row and every column.

Next we prove (g) $\Leftrightarrow$ (f). If T_A is an isomorphism, then it must be injective by definition. To prove the converse, assume T_A is injective, then by Theorem 6.3, $\text{rref}(A)$ has a pivot in every column. Since A has n columns, then $\text{rref}(A)$ have n pivots. But it also has n rows, so $\text{rref}(A)$ has a pivot in every row. By Theorem 6.3, T_A is surjective. Since T_A is injective and surjective, it is, by definition, bijective. Therefore, T_A is an isomorphism.

Next we prove (c) $\Leftrightarrow$ (g). Suppose T_A is injective. By an exercise in chapter 5, we know that $T_A(\vec{0}) = \vec{0}$. By the definition of injectivity, no other vector maps to $\vec{0}$, and so we must have $\ker(T_A) = \{\vec{0}\}$. By lecture activity 5.8, we know that $\text{Nul}(A) = \ker(T_A)$. It follows that if T_A is injective, then $\text{Nul}(A) = \{\vec{0}\}$. Now to prove the converse, suppose that $\text{Nul}(A) = \ker(T_A) = \{\vec{0}\}$, and that $T_A(\vec{x}) = T_A(\vec{y})$ for some vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then

$$\begin{aligned} \vec{0} &= T_A(\vec{x}) - T_A(\vec{y}) \\ &= T_A(\vec{x} - \vec{y}) && \text{by the linearity of } T_A. \end{aligned}$$

It follows that $\vec{x} - \vec{y} \in \ker(T_A) = \{\vec{0}\}$. Therefore, $\vec{x} - \vec{y} = \vec{0}$, or $\vec{x} = \vec{y}$, which shows that T_A is injective.

Next we prove (d) $\Leftrightarrow$ (h). But this is by the definition of surjectivity. Note that $T_A : \mathbb{R}^n \to \mathbb{R}^n$, and so T_A is surjective if and only if $\operatorname{im}(T_A) = \mathbb{R}^n$. But be lecture activity 5.8, $\operatorname{im}(T_A) = \operatorname{Col}(A)$. It follows that T_A is surjective if and only if $\operatorname{Col}(A) = \mathbb{R}^n$.

C.7. Chapter 7 Exercise Solutions

P7.1 Let E be an $n \times n$ elementary matrix, and let B be any $n \times n$ matrix. We need to show that $\det(EB) = \det(E)\det(B)$. Now there are three types of elementary matrices, let's consider each one.

(a) Let E be an elementary matrix that is obtained by interchanging two rows of the identity matrix, then by Theorem 7.20.1, we get that $\det(E) = -\det(I_n) = -1$. Note that the matrix EB is the matrix obtained from the matrix B by interchanging two rows of B. And so by Theorem 7.20.1, we can see that $\det(EB) = -\det(B)$, and since $\det(E) = -1$, it follows that $\det(EB) = \det(E)\det(B)$.

(b) Next, let E be an elementary matrix that is obtained by multiplying one row of the identity matrix by a scalar c. By Theorem 7.20.2, we have $\det(E) = c\det(I_n) = c \cdot 1 = c$. Note that the matrix EB is the matrix obtained by multiplying a row of B by a scalar c. So, by Theorem 7.20.2, we can conclude that $\det(EB) = c\det(B) = \det(E)\det(B)$.

(c) Finally, let E be an elementary matrix that is obtained by replacing a row of the identity matrix by the same row added to a scalar multiple of another row of the identity matrix. By Theorem 7.20.3, we have $\det(E) = \det(I_n) = 1$. Now the matrix EB is obtained by replacing a row of the B by the same row added to a scalar multiple of another row of B. So, by Theorem 7.20.3, we have $\det(EB) = \det(B) = 1 \cdot \det(B) = \det(E)\det(B)$.

P7.2 (a) Let A and B be $n \times n$ matrices. Let $A = [a_{ij}], B = [b_{ij}]$, and let $AB = [c_{ij}]$. Then $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$ and $(AB)^{\top}_{ij} = c_{ji} = \sum_{k=1}^{n} a_{jk}b_{ki}$. Now $(A^{\top})_{ij} = a_{ji}$ and $(B^{\top})_{ij} = b_{ji}$. If $B^{\top}A^{\top} = [d_{ij}]$, then $d_{ij} = \sum_{k=1}^{n} a_{jk}b_{ki} = c_{ji}$. Therefore, we have $(AB)^{\top} = B^{\top}A^{\top}$.

(b) Let E be an $n \times n$ elementary matrix. We have three cases:

(i) E is obtained by interchanging R_i and R_j of the identity matrix I_n. Then $E^T = E$, and so they must have the same determinant.

(ii) E is obtained from the identity matrix I_n by multiplying R_i by a scalar c. Then $E^T = E$, and so they must have the same determinant.

(iii) E is obtained from the identity matrix I_n by replacing R_i with $R_i + kR_j$ for some scalar k. Then E^T will be the elementary matrix obtained from the identity matrix by replacing R_j with $R_j + kR_i$. As we saw in exercise 1, we have $\det(E) = 1$, and $\det(E^{\top}) = 1$. And so $\det(E) = \det(E^T)$.

P7.3 (a) Let A be an $n \times n$ matrix, and suppose A is not invertible. For the sake of contradiction, suppose $A^{\top}$ is invertible. Then there exists a unique matrix B such that $A^{\top}B = I_n$. Transposing both sides of that equation gives us

$$\begin{aligned} (A^\top B)^\top &= I_n^\top \\ B^\top (A^\top)^\top &= I_n && \text{using the first part of Exercise 2} \\ B^\top A &= I_n. \end{aligned}$$

But this implies that A is invertible with $A^{-1} = B^\top$, which is a contradiction. Therefore, $A^\top$ is not invertible and by Corollary 7.21, $\det(A^\top) = 0$.

(b) Let A and B be $n \times n$ matrices. Assume A is not invertible, then by Corollary 7.21 we have $\det(A) = 0$. For the sake of contradiction, suppose AB is invertible. Then there exists a matrix C such that $ABC = I_n$. But then $A(BC) = I_n$, and BC is the inverse of A, which contradicts the non-invertibility of A. So we must have that AB is not invertible, and by Corollary 7.21, we have $\det(AB) = 0$. In this case,

$$\det(AB) = 0 = 0 \cdot \det(B) = \det(A)\det(B).$$

Similarly, if we let A and B be $n \times n$ matrices, with B non-invertible, we will have $\det(B) = 0$. Suppose AB is invertible, then there exists a matrix C such that $CAB = I_n$. But then $(CA)B = I_n$, and CA would be the inverse of B. This contradicts the non-invertibility of B. It follows that AB is not invertible, and by Corollary 7.21, we have $\det(AB) = 0$. In this case,

$$\det(AB) = 0 = \det(A) \cdot 0 = \det(A)\det(B).$$

P7.4 Let A be an invertible matrix. Then $\det(A) \neq 0$ by Theorem 7.18. By Proposition 7.23, we have

$$\det(AA^{-1}) = \det(A)\det(A^{-1}).$$

But $\det(AA^{-1}) = \det(I_n) = 1$, which give us $1 = \det(A)\det(A^{-1})$. Since $\det(A) \neq 0$, we may divide by it to get $\det(A^{-1}) = \dfrac{1}{\det(A)}$.

P7.5 The statement is false. Consider $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Note that $\det(I_2) = 1$. Now $-I_2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, and so $\det(-I_2) = (-1)(-1) - 0 \cdot 0 = 1 = \det(I_2)$. So we have found a matrix $A = I_2$ for which $\det(-A) \neq -\det(A)$.

P7.6 The statement is false. Consider the matrices $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, and $B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. We have $A + B = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = I_2$. Then $\det(A) = \det(B) = 0$, so that $\det(A) + \det(B) = 0$, but $\det(A + B) = 1$.

P7.7 The statement is True. Let A and B be $n \times n$ matrices and assume AB is invertible. Then by Theorem 7.18 we have $\det(AB) \neq 0$. By Proposition 7.23, we have $\det(AB) = \det(A)\det(B) \neq 0$. But this implies that $\det(A) \neq 0$ and $\det(B) \neq 0$. By Theorem 7.18, it follows that both A and B are invertible.

P7.8 Let A and B be $n \times n$ matrices with $\operatorname{rank}(A) = \operatorname{rank}(B) = n$. Then $\operatorname{rref}(A) = \operatorname{rref}(B) = I_n$ and by the Invertible Matrix Theorem (Theorem 6.21), A and

B are invertible. By Theorem 7.18 we have $\det(A) \neq 0$ and $\det(B) \neq 0$. Now $\det(AB) = \det(A)\det(B)$ by Proposition 7.23, and so $\det(AB) \neq 0$. It follows that AB is invertible (by Theorem 7.18). Applying the Invertible Matrix Theorem again, we can see that $\mathrm{rref}(AB) = I_n$, and so $\mathrm{rank}(AB) = n$.

P7.9 First, let's prove the following claim: For an $n \times n$ matrix A, we have $\det(A) = (-1)^n \det(A)$. Note that $-A = E_n E_{n-1} \cdots E_2 E_1 A$, where E_i is the elementary matrix obtained by multiplying the ith row of the identity matrix I_n by -1. By repeated application of Proposition 7.23, we get that $\det(-A) = \det(E_n) \cdots \det(E_2)\det(E_1)\det(A)$. Now E_i is the elementary matrix obtained by multiplying the ith row of the identity matrix I_n by -1, and so by Theorem 7.20 we have that $\det(E_i) = -\det(I_n) = -1$. This gives us

$$\det(-A) = \det(E_n) \cdots \det(E_2)\det(E_1)\det(A) = (-1)^n \det(A).$$

Now we're ready to prove the original statement. Let A be an $n \times n$ skew-symmetric matrix, where n is an odd integer. Now

$$\begin{aligned}
\det(A) &= \det(-A^\top) \\
&= (-1)^n \det(A^\top) && \text{by our claim above} \\
&= -\det(A^\top) && \text{since } n \text{ is odd} \\
&= -\det(A) && \text{by Proposition 7.23.}
\end{aligned}$$

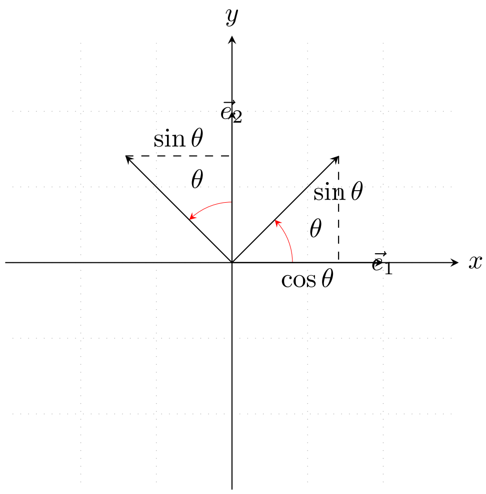
But this implies that $\det(A) = 0$, and so by Theorem 7.18, A is not invertible.

P7.10 Let A be an $n \times n$ be a nilpotent matrix with $A^m = 0$ for a positive integer m. Suppose A is invertible. Then $\det(A) \neq 0$ by Theorem 7.18. By repeatedly applying Proposition 7.23, we get $\det(A^m) = \det(A) \cdots \det(A) = \det(A)^m$. Since $\det(A) \neq 0$, it follows that $\det(A)^m \neq 0$, and consequently $\det(A^m) \neq 0$. But this is a contradiction, because the zero matrix has determinant equal to zero. Therefore, it must be the case that $\det(A) = 0$, and A is not invertible.

C.8. Chapter 8 Exercise Solutions

P8.1 (a) Recall that the columns of the matrix A_F are the transformed standard basis vectors, $F(\vec{e}_1)$ and $F(\vec{e}_2)$. Rotating the vector $\vec{e}_1$ counterclockwise by an angle of θ, you will get the vector $\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ as you can see in the figure (using basic trigonometry). And rotating the vector $\vec{e}_2$

counterclockwise by an angle of θ, you will get the vector $\begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$.



This gives us the associated matrix for F, which is $A_F = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

(b) To find the eigenvalues of A_F, we start by computing its characteristic polynomial,

$$\begin{aligned}
\chi_{A_F}(x) &= \det(A_F - xI) \\
&= \det \begin{pmatrix} \cos\theta - x & -\sin\theta \\ \sin\theta & \cos\theta - x \end{pmatrix} \\
&= (\cos\theta - x)^2 + \sin^2\theta \\
&= \cos^2\theta - 2x\cos\theta + x^2 + \sin^2\theta \\
&= x^2 - 2x\cos\theta + 1 && \text{since } cos^2\theta + \sin^2\theta = 1.
\end{aligned}$$

Solving the equation $x^2 - 2x\cos\theta + 1 = 0$, we find that $x = \cos\theta \pm \sqrt{\cos^2\theta - 1}$. But $\cos^2\theta - 1 < 0$ whenever $\cos\theta \neq 0$. Therefore, A_F only has real eigenvalues when $\cos\theta = 0$. And so A_F only has real eigenvalues when θ is an integer multiple of $180°$.

P8.2 Let A be an $n \times n$ matrix. Then

$$\begin{aligned}
\chi_A(x) &= \det(A - xI) \\
&= \det(A - xI)^\top && \text{by Proposition 7.23} \\
&= \det(A^\top - xI^\top) \\
&= \det(A^\top - xI) && \text{since } I^\top = I \\
&= \chi_{A^\top}(x).
\end{aligned}$$

P8.3 The statement is false. Consider the matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. The vector $\vec{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is an eigenvector of A, since $A\vec{v} = \vec{v}$. However,

$$A^\top \vec{v} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \neq c \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

for any scalar c. This means that $\vec{v}$ is not an eigenvector for $A^\top$.

P8.4 Note that the eigenvalues of a matrix A are the roots of its characteristic polynomial. So, we have

$$\begin{aligned} \chi_A(x) &= \det(A - xI) \\ &= \underbrace{(\lambda_1 - x) \cdots (\lambda_1 - x)}_{m_1 \text{ times}} \cdots \underbrace{(\lambda_k - x) \cdots (\lambda_k - x)}_{m_k \text{ times}} \end{aligned}$$

Now if we set $x = 0$, then

$$\chi_A(0) = \det(A) = \underbrace{\lambda_1 \cdots \lambda_1}_{m_1 \text{ times}} \cdots \underbrace{\lambda_k \cdots \lambda_k}_{m_k \text{ times}} = \lambda_1^{m_1} \cdots \lambda_k^{m_k}.$$

P8.5 Let A be an $n \times n$ matrix. Suppose that A is invertible. Then $\det(A) \neq 0$ by Theorem 7.18. But this means that $\chi_A(0) = \det(A - 0 \cdot I) = \det(A) \neq 0$, and so 0 is not a root of the characteristic polynomial. It follows that 0 is not an eigenvalue of A. For the converse, suppose A is not invertible. Then $\det(A) = 0$ by Theorem 7.18. And we have $\chi_A(0) = \det(A - 0 \cdot I) = \det(A) = 0$, and so 0 is a root of the characteristic polynomial. In other words, 0 is an eigenvalue of A.

P8.6 (a) If $A = I_n$, then A is clearly invertible. For the converse, let A be an $n \times n$ idempotent matrix. Then $A^2 = A$, and we have,

$$A = A^{-1}A^2 = A^{-1}A = I_n.$$

(b) Let A be an idempotent matrix that's not equal to the identity matrix. Suppose λ is an eigenvalue of A with eigenvector $\vec{v}$. Then $A\vec{v} = \lambda v$. We have

$$\lambda\vec{v} = A\vec{v} = A^2\vec{v} = A(A\vec{v}) = A(\lambda\vec{v}) = \lambda(A\vec{v}) = \lambda(\lambda\vec{v}) = \lambda^2\vec{v}.$$

It follows that $(\lambda^2 - \lambda)\vec{v} = \vec{0}$. Since $\vec{v}$ is an eigenvector, $\vec{v} \neq 0$. So it must be the case that $\lambda^2 - \lambda = \lambda(\lambda - 1) = 0$. It follows that $\lambda = 0$ or $\lambda = 1$.

C.9. Chapter 9 Exercise Solutions

P9.1 The statement is false. The set $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$ is a generating set for $\mathbb{R}^2$, but the vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ has multiple representations in terms of $\mathcal{B}$. For example, $\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} + 0 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 0 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

P9.2 Let $\mathcal{C} = \{\vec{c}_1, \cdots, \vec{c}_n\}$ be a basis for a vector space V, and let $\vec{x}, \vec{y} \in V$ and scalar $k \in \mathcal{R}$. Then, by Theorem 9.2, there are unique real numbers $x_1, \cdots, x_n, y_1, \cdots, y_n$ such that

$$\vec{x} = x_1\vec{c}_1 + \cdots + x_n\vec{c}_n, \text{ and } \vec{y} = y_1\vec{c}_1 + \cdots + y_n\vec{c}_n.$$

So, we have $[\vec{x}]_{\mathcal{C}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, and $[\vec{y}]_{\mathcal{C}} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$. Now $\vec{x} + \vec{y} = x_1\vec{c}_1 + \cdots + x_n\vec{c}_n + y_1\vec{c}_1 + \cdots + y_n\vec{c}_n = (x_1 + y_1)\vec{c}_1 + \cdots + (x_n + y_n)\vec{c}_n$. So,

$$[\vec{x} + \vec{y}]_{\mathcal{C}} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = [\vec{x}]_{\mathcal{C}} + [\vec{y}]_{\mathcal{C}}.$$

Additionally, $k\vec{x} = k(x_1\vec{c}_1 + \cdots + x_n\vec{c}_n) = kx_1\vec{c}_1 + \cdots + kx_n\vec{c}_n$. So,

$$[k\vec{x}]_{\mathcal{C}} = \begin{pmatrix} kx_1 \\ \vdots \\ kx_n \end{pmatrix} = k\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = k[\vec{x}]_{\mathcal{C}}.$$

P9.3 Let V be a vector space with basis $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_n\}$, and let $\mathcal{C}$ be any other basis of V. We want to show that the set $S = \{[\vec{b}_1]_{\mathcal{C}}, \ldots, [\vec{b}_n]_{\mathcal{C}}\}$ is linearly independent. Note that $M_{\mathcal{C}\leftarrow\mathcal{B}}$ defines an injective linear transformation (since each vector in V is represented uniquely as a linear combination of the vectors in $\mathcal{C}$). Since the $\mathcal{B}$ is a basis for V, it's vectors are linearly independent, and by Exercise 5.11, it follows that the images of these vectors under $M_{\mathcal{C}\leftarrow\mathcal{B}}$ are also linearly independent. In other words, the set $S = \{[\vec{b}_1]_{\mathcal{C}}, \ldots, [\vec{b}_n]_{\mathcal{C}}\}$ is linearly independent.

P9.4 The change of basis matrix $M_{\mathcal{C}\leftarrow\mathcal{B}}$ will be 3×3. Let $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \vec{b}_3\}$, and $\mathcal{C} = \vec{c}_1, \vec{c}_2, \vec{c}_3$. For a vector $\vec{v} \in V$, there are unique real numbers $x_1, x_2, x_3, y_1, y_2, y_3$ such that $\vec{v} = x_1\vec{b}_1 + x_2\vec{b}_2 + x_3\vec{b}_3$, and $\vec{v} = y_1\vec{c}_1 + y_2\vec{c}_2 + y_3\vec{c}_3$. So, $[\vec{v}]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, and $[\vec{v}]_{\mathcal{C}} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$. Now

$$M_{\mathcal{C}\leftarrow\mathcal{B}}[\vec{v}]_{\mathcal{B}} = [\vec{v}]_{\mathcal{C}}.$$

Since $[\vec{v}]_{\mathcal{B}}$ and $[\vec{v}]_{\mathcal{C}}$ are both elements of $\mathbb{R}^3$, it follows that $M_{\mathcal{C}\leftarrow\mathcal{B}}$ has to be 3×3.

P9.5 Let $\mathcal{B}, \mathcal{C}$ and $\mathcal{D}$ be bases for a vector space V. Let $\vec{v} \in V$, then

$$M_{\mathcal{C}\leftarrow\mathcal{B}}M_{\mathcal{B}\leftarrow\mathcal{D}}[\vec{v}]_{\mathcal{D}} = M_{\mathcal{C}\leftarrow\mathcal{B}}[\vec{v}]_{\mathcal{B}} = [\vec{v}]_{\mathcal{C}}.$$

It follows that $M_{\mathcal{C}\leftarrow\mathcal{B}}M_{\mathcal{B}\leftarrow\mathcal{D}} = M_{\mathcal{C}\leftarrow\mathcal{D}}$.

C.10. Chapter 10 Exercise Solutions

P10.1 (a) Let λ_1 and λ_2 be distinct eigenvalues of a matrix A, and suppose that $\vec{v}_1 \in E_{\lambda_1}$ and $\vec{v}_2 \in E_{\lambda_2}$. Suppose that $x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{0}$. Then

$$\begin{aligned} A(x_1\vec{v}_1 + x_2\vec{x}_2) &= A\vec{0} \\ x_1A\vec{v}_1 + x_2A\vec{v}_2 &= \vec{0} \\ x_1\lambda_1\vec{v}_1 + x_2\lambda_2\vec{v}_2 &= \vec{0}. \end{aligned}$$

Also, multiplying the equation $x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{0}$ by λ_1 results in $x_1\lambda_1\vec{v}_1 + x_2\lambda_1\vec{v}_2 = \vec{0}$.

(b) Taking the difference of the two equalities we've show above, we get

$$\begin{aligned} x_1\lambda_1\vec{v}_1 + x_2\lambda_2\vec{v}_2 - x_1\lambda_1\vec{v}_1 - x_2\lambda_1\vec{v}_2 &= \vec{0} - \vec{0} \\ x_2\lambda_2\vec{v}_2 - x_2\lambda_1\vec{v}_2 &= \vec{0} \\ x_2(\lambda_1 - \lambda_2)\vec{v}_2 &= \vec{0}. \end{aligned}$$

Since λ_1, and λ_2 are distinct, $\lambda_1 - \lambda_2 \neq 0$. And since $\vec{v}_2$ is an eigenvector, $\vec{v}_2 \neq \vec{0}$. It follows that $x_2 = 0$.

(c) Similarly, if we multiply $x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{0}$ by λ_2, we get $x_1\lambda_2\vec{v}_1 + x_2\lambda_2\vec{x}_2 = \vec{0}$. Subtracting this from $x_1\lambda_1\vec{v}_1 + x_2\lambda_2\vec{v}_2 = \vec{0}$, which we showed is true in (a), we get

$$\begin{aligned} x_1\lambda_1\vec{v}_1 + x_2\lambda_2\vec{v}_2 - x_1\lambda_2\vec{v}_1 - x_2\lambda_2\vec{v}_2 &= \vec{0} \\ x_1\lambda_1\vec{v}_1 - x_1\lambda_2\vec{v}_1 &= \vec{0} \\ x_1(\lambda_1 - \lambda_2)\vec{v}_1 &= \vec{0}. \end{aligned}$$

Since λ_1, and λ_2 are distinct, $\lambda_1 - \lambda_2 \neq 0$. And since $\vec{v}_1$ is an eigenvector, $\vec{v}_1 \neq \vec{0}$. It follows that $x_1 = 0$. Since $x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{0}$ only has the trivial solution, it follows that $\vec{v}_1$ and $\vec{v}_2$ are linearly independent.

(d) Suppose that $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are eigenvectors of a matrix A with corresponding eigenvalues $\lambda_1, \lambda_2, \lambda_3$, where the eigenvalues are all distinct. Suppose that $x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 = \vec{0}$. Then

$$\begin{aligned} A(x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3) &= A\vec{0} \\ x_1\lambda_1\vec{v}_1 + x_2\lambda_2\vec{v}_2 + x_3\lambda_3\vec{v}_3 &= \vec{0}. \end{aligned}$$

We also have $x_1\lambda_1\vec{v}_1 + x_2\lambda_1\vec{v}_2 + x_3\lambda_1\vec{v}_3 = \vec{0}$, Subtracting these two equations, we get

$$\begin{aligned} x_1\lambda_1\vec{v}_1 + x_2\lambda_2\vec{v}_2 + x_3\lambda_3\vec{v}_3 - x_1\lambda_1\vec{v}_1 - x_2\lambda_1\vec{v}_2 - x_3\lambda_1\vec{v}_3 &= \vec{0} \\ x_2(\lambda_2 - \lambda_1)\vec{v}_2 + x_3(\lambda_3 - \lambda_1)\vec{v}_2 &= \vec{0}. \end{aligned}$$

By our result in the previous parts, we must have $x_2(\lambda_2 - \lambda_1) = 0$ and $x_3(\lambda_3 - \lambda_1) = 0$. Since the eigenvalues are distinct, $\lambda_2 - \lambda_1 \neq 0$ and $\lambda_3 - \lambda_1 \neq 0$. It follows that $x_2 = x_3 = 0$. But then, $x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 = \vec{0}$ reduces to $x_1\vec{v}_1 = \vec{0}$, and since $\vec{v}_1 \neq \vec{0}$, we have $x_1 = 0$. Hence, $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a linearly independent set.

(e) Suppose that $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are eigenvectors of a matrix A with corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, where the eigenvalues are all distinct. Using the same strategy as above, we can show that if $x_1\vec{v}_1 + \cdots + x_n\vec{v}_n = \vec{0}$, then $x_2(\lambda_2 - \lambda_1)\vec{v}_2 + \cdots + x_n(\lambda_n - \lambda_1)\vec{v}_n = \vec{0}$, reducing it to a problem concerning $n-1$ vectors. Continuing in this way, we can keep reducing it until we reach a problem concerning 3 vectors and use our result above to show that $x_n = x_{n-1} = x_{n-2} = 0$ and back substitute to get that all the scalars x_i must be equal to zero. This then shows that the set of vectors $\{\vec{v}_1, \dots, \vec{v}_n\}$ is linearly independent.

P10.2 Suppose A and B are similar matrices, then there exists an invertible matrix P such that $A = P^{-1}BP$. Then

$$\begin{aligned}
\chi_A(x) &= \det(A - xI) \\
&= \det(P^{-1}BP - xI) \\
&= \det(P^{-1}(B - xI)P) && \text{since } P^{-1}(xI)P = x(P^{-1}IP) = xI \\
&= \det(P^{-1})\det(B - xI)\det(P) \\
&= \det(B - xI) && \text{since } \det(P^{-1}) = \frac{1}{\det(P)} \\
&= \chi_B(x).
\end{aligned}$$

P10.3 Suppose A is diagonalizable. Then there exists a diagonal matrix D and an invertible matrix P such that $A = PDP^{-1}$. By Proposition **??**, for any positive integer n, we have $A^n = PD^nC^{-1}$, showing that A^n is also diagonalizable.

P10.4 Suppose A is diagonalizable. Then there exists a diagonal matrix D and an invertible matrix P such that $A = PDP^{-1}$. Taking the transpose of both sides, we get

$$\begin{aligned}
A^\top &= (PDP^{-1})^\top \\
&= (P^{-1})^\top D^\top P^\top \\
&= (P^\top)^{-1} D^\top P^\top \\
&= CD^\top C^{-1},
\end{aligned}$$

where $C = (P^\top)^{-1}$. Therefore, $P^\top$ is diagonalizable.

P10.5 The statement is true. Let A be a nilpotent diagonalizable matrix. Then there is a positive integer m such that A^m is equal to the zero matrix. Additionally, there is a diagonal matrix D and an invertible matrix P such that $A = PDP^{-1}$. Now $A^m = PD^mP^{-1}$ is equal to the zero matrix. But this implies that D is equal to the zero matrix, and therefore, A is equal to the zero matrix.

P10.6 The statement is false. The zero matrix is diagonalizable (it is already a diagonal matrix), but it is not invertible.

P10.7 Note that $A_F = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$, and

$$\chi_{A_F}(x) = \det(A_F - xI) = \det \begin{pmatrix} 1-x & m \\ 0 & 1-x \end{pmatrix} = (1-x)^2.$$

Therefore, A_F has only one eigenvalue, $\lambda = 1$. Now to find the dimension of its eigenspace. Observe that $A_F - I = \begin{pmatrix} 0 & m \\ 0 & 0 \end{pmatrix}$ and if $m \neq 0$, $A_F - I$ has nullity equal to 1. In other words, if $m \neq 0$, then λ has algebraic multiplicity equal to 2 but its geometric multiplicity is equal to 1. By the Diagonalization Theorem, F is not diagonalizable if $m \neq 0$.

Similarly, $A_G = \begin{pmatrix} 1 & 0 \\ m & 1 \end{pmatrix}$, and

$$\chi_{A_G}(x) = \det(A_G - xI) = \det \begin{pmatrix} 1-x & 0 \\ m & 1-x \end{pmatrix} = (1-x)^2.$$

Therefore, A_G has only one eigenvalue, $\lambda = 1$. Now to find the dimension of its eigenspace. Observe that $A_G - I = \begin{pmatrix} 0 & 0 \\ m & 0 \end{pmatrix}$ and if $m \neq 0$, $A_G - I$ has nullity equal to 1. In other words, if $m \neq 0$, then λ has algebraic multiplicity equal to 2 but its geometric multiplicity is equal to 1. By the Diagonalization Theorem, G is not diagonalizable if $m \neq 0$.

P10.8 (a) We have to find f_i for $i \in \{1, \ldots, 12\}$. Using the recurrence relation, we have

$$\begin{aligned}
f_2 &= f_1 + f_0 &&= 1 + 0 = 1 \\
f_3 &= f_2 + f_1 &&= 1 + 1 = 2 \\
f_4 &= f_3 + f_2 &&= 2 + 1 = 3 \\
f_5 &= f_4 + f_3 &&= 3 + 2 = 5 \\
f_6 &= f_5 + f_4 &&= 5 + 3 = 8 \\
f_7 &= f_6 + f_5 &&= 8 + 5 = 13 \\
f_8 &= f_7 + f_6 &&= 13 + 8 = 21 \\
f_9 &= f_8 + f_7 &&= 21 + 13 = 34 \\
f_{10} &= f_9 + f_8 &&= 34 + 21 = 55 \\
f_{11} &= f_{10} + f_9 &&= 55 + 34 = 89 \\
f_{12} &= f_{11} + f_{10} &&= 89 + 55 = 144.
\end{aligned}$$

(b) Note that $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} f_{n-1} \\ f_n \end{pmatrix} = \begin{pmatrix} f_n \\ f_{n-1} + f_n \end{pmatrix} = \begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix}$.

(c) Observe that

$$\begin{aligned}
\begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} f_{n-1} \\ f_n \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} f_{n-2} \\ f_{n-1} \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} f_{n-3} \\ f_{n-2} \end{pmatrix} \\
&\vdots \\
&= \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}}_{n \text{ times}} \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^n \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix}.
\end{aligned}$$

(d) Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, then

$$\chi_A(x) = \det(A - xI) = \det \begin{pmatrix} -x & 1 \\ 1 & 1-x \end{pmatrix} = -x(1-x) - 1 = x^2 - x - 1.$$

The eigenvalues of A are the roots of $\chi_A(x)$, which are $\alpha = \frac{1+\sqrt{5}}{2}$ and $\beta = \frac{1-\sqrt{5}}{2}$. Next, we find the eigenvectors associated to each eigenvalue. For α, we have $A - \alpha I = \begin{pmatrix} -\alpha & 1 \\ 1 & 1-\alpha \end{pmatrix}$ which can be row reduced to $\begin{pmatrix} -\alpha & 1 \\ 0 & 0 \end{pmatrix}$. So the α-eigenspace is equal to $\operatorname{Span}\left\{\begin{pmatrix} 1 \\ \alpha \end{pmatrix}\right\}$. For β, we have $A - \beta I = \begin{pmatrix} -\beta & 1 \\ 1 & 1-\beta \end{pmatrix}$ which can be row reduced to $\begin{pmatrix} -\beta & 1 \\ 0 & 0 \end{pmatrix}$. So the β-eigenspace is equal to $\operatorname{Span}\left\{\begin{pmatrix} 1 \\ \beta \end{pmatrix}\right\}$. This gives us the following eigendecomposition of A

$$A = PDP^{-1},$$

where $D = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$, and $P = \begin{pmatrix} 1 & 1 \\ \alpha & \beta \end{pmatrix}$. Note that $P^{-1} = \frac{1}{\beta-\alpha} \begin{pmatrix} \beta & -1 \\ -\alpha & 1 \end{pmatrix}$. Now using the previous part,

$$\begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix} = A^n \begin{pmatrix} 0 \\ 1 \end{pmatrix} = PD^nP^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

So we have

$$\begin{aligned}
\begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ \alpha & \beta \end{pmatrix} \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}^n \begin{pmatrix} \frac{\beta}{\beta-\alpha} & \frac{-1}{\beta-\alpha} \\ \frac{-\alpha}{\beta-\alpha} & \frac{1}{\beta-\alpha} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 1 \\ \alpha & \beta \end{pmatrix} \begin{pmatrix} \alpha^n & 0 \\ 0 & \beta^n \end{pmatrix} \begin{pmatrix} \frac{\beta}{\beta-\alpha} & \frac{-1}{\beta-\alpha} \\ \frac{-\alpha}{\beta-\alpha} & \frac{1}{\beta-\alpha} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} \alpha^n & \beta^n \\ \alpha^{n+1} & \beta^{n+1} \end{pmatrix} \begin{pmatrix} \frac{\beta}{\beta-\alpha} & \frac{-1}{\beta-\alpha} \\ \frac{-\alpha}{\beta-\alpha} & \frac{1}{\beta-\alpha} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{-\alpha^n\beta-\alpha\beta^n}{\beta-\alpha} & \frac{\beta^n-\alpha^n}{\beta-\alpha} \\ \frac{-\beta\alpha^{n+1}-\alpha\beta^{n+1}}{\beta-\alpha} & \frac{\beta^{n+1}-\alpha^{n+1}}{\beta-\alpha} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{\beta^n-\alpha^n}{\beta-\alpha} \\ \frac{\beta^{n+1}-\alpha^{n+1}}{\beta-\alpha} \end{pmatrix}.
\end{aligned}$$

And so, we get $f_n = \frac{\beta^n-\alpha^n}{\beta-\alpha} = \frac{\alpha^n-\beta^n}{\alpha-\beta}$.

(e) Using the previous part, we can compute f_{12} using the formula

$$f_{12} = \frac{\alpha^{12}-\beta^{12}}{\alpha-\beta} = 144,$$

which agrees with what we found above.
Also, $f_{20} = \frac{\alpha^{20}-\beta^{20}}{\alpha-\beta} = 6765$, $f_{50} = \frac{\alpha^{50}-\beta^{50}}{\alpha-\beta} = 12,586,269,025$, and $f_{100} = \frac{\alpha^{100}-\beta^{100}}{\alpha-\beta} = 354,224,848,179,261,915,075$.

C.11. Chapter 11 Exercise Solutions

P11.1 The statement is true. Let $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$. Then $\vec{x}\cdot\vec{x} = x_1^2 + \cdots + x_n^2$. If $\vec{x}\cdot\vec{x} = x_1^2+\cdots+x_n^2 = 0$, then since $x_i^2 \geq 0$ it must be the case that $x_1 = \cdots = x_n = 0$. In other words, we have $\vec{x} = \vec{0}$.

P11.2 The statement is false. Consider the vectors $\vec{x} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ and $\vec{y} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. We have $\vec{x}\cdot\vec{y} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = (-1)(1) + (0)(0) = -1 < 0$.

P11.3 The statement is true. Let $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$. Then $\|\vec{x}\| = \sqrt{\vec{x}\cdot\vec{x}} = \sqrt{x_1^2+\cdots+x_n^2}$. Since $x_i^2 \geq 0$ for each $i \in \{1,\ldots,n\}$, it follows that $x_1^2 + \cdots + x_n^2 \geq 0$ and $\|\vec{x}\| \geq 0$.

P11.4 The statement is true. We showed in a previous exercise that if $\vec{x}$ is a vector in $\mathbb{R}^n$ and $\vec{x}\cdot\vec{x} = 0$ then $\vec{x} = \vec{0}$. To show the converse, suppose that $\vec{x} = \vec{0}$, then $\vec{x} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$. In this case,

$$\vec{x} \cdot \vec{x} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = 0 + \cdots + 0 = 0.$$

P11.5 The statement is false. Consider the vectors $\vec{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $\vec{y} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. We have $\|\vec{x}\| = \sqrt{2^2 + 1^2} = \sqrt{5}$, and $\|\vec{y}\| = \sqrt{1^2 + 2^2} = \sqrt{5}$. So that $\|\vec{x}\| + \|\vec{y}\| = \sqrt{5} + \sqrt{5} = 2\sqrt{5}$. On the other hand, $\vec{x} + \vec{y} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$, so that

$$\|\vec{x} + \vec{y}\| = \sqrt{3^2 + 3^2} = 2\sqrt{3} \neq 2\sqrt{5} = \|\vec{x}\| + \|\vec{y}\|.$$

P11.6 Let $\vec{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$, $\vec{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ and $\vec{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{R}^n$, and let c be a scalar.

(a) Commutativity: Note that

$$\begin{aligned} \vec{u} \cdot \vec{v} &= u_1 v_1 + \cdots + u_n v_n \\ &= v_1 u_1 + \cdots v_n u_n \qquad \text{since multiplication in } \mathbb{R} \text{ is commutative} = \vec{v} \cdot \vec{u}. \end{aligned}$$

(b) Distributivity with addition: Note that

$$\begin{aligned} (\vec{u} + \vec{v}) \cdot \vec{w} &= \begin{pmatrix} u_1 + v_1 \\ \vdots \\ u_n + v_n \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \\ &= (u_1 + v_1) w_1 + \cdots + (u_n + v_n) w_n \\ &= u_1 w_1 + v_1 w_1 + \cdots + u_n w_n + v_n w_n \qquad \text{by the distributivity of multiplication over addition} \\ &= (u_1 w_1 + \cdots + u_n w_n) + (v_1 w_1 + \cdots + v_n w_n) \\ &= \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}. \end{aligned}$$

(c) Distributivity with Scalar Multiplication: Note that

$$\begin{aligned} (c\vec{u}) \cdot \vec{v} &= \begin{pmatrix} cu_1 \\ \vdots \\ cu_n \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \\ &= cu_1 v_1 + \cdots + cu_n v_n \\ &= c(u_1 v_1 + \cdots + u_n v_n) \\ &= c(\vec{u} \cdot \vec{v}). \end{aligned}$$

P11.7 Let $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$. Then

$$\begin{aligned} d(\vec{x}, \vec{y}) &= \|\vec{x} - \vec{y}\| \\ &= \sqrt{(\vec{x} - \vec{y}) \cdot (\vec{x} - \vec{y})} \\ &= \sqrt{\vec{x} \cdot \vec{x} - \vec{x} \cdot \vec{y} - \vec{y} \cdot \vec{x} + \vec{y} \cdot \vec{y}} \\ &= \sqrt{\vec{y} \cdot \vec{y} - \vec{y} \cdot \vec{x} - \vec{x} \cdot \vec{y} + \vec{x} \cdot \vec{x}} \\ &= \sqrt{(\vec{y} - \vec{x}) \cdot (\vec{y} - \vec{x})} \\ &= \|\vec{y} - \vec{x}\| \end{aligned}$$

P11.8 Let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a set of nonzero orthogonal vectors in $\mathbb{R}^n$. We will show that $\mathcal{B}$ is linearly independent, and since it contains n vectors, it will follow that $\mathcal{B}$ is a basis for $\mathbb{R}^n$. Consider the vector equation

$$a_1\vec{v}_1 + \cdots + a_n\vec{v}_n = \vec{0}.$$

Take the dot product of both sides of this equation with $\vec{v}_i$, where $i \in \{1, \cdots, n\}$. We will get

$$a_1\vec{v}_1 \cdot \vec{v}_i + \cdots + a_n\vec{v}_n \cdot \vec{v}_i = \vec{0} \cdot \vec{v}_i = 0.$$

Since the vectors are orthogonal, $\vec{v}_j \cdot \vec{v}_i = 0$ whenever $i \neq j$, and the last equation reduces to

$$a_i\vec{v}_i \cdot \vec{v}_i = 0.$$

In other words, we have $a_i\|\vec{v}_i\|^2 = 0$. We assumed that $\vec{v}_i$ is a nonzero vector, and so by a previous exercise, $\|\vec{v}_i\|^2 \neq 0$. It follows that $a_i = 0$ for each $i \in \{1, \cdots, n\}$, and therefore, the set $\mathcal{B}$ is linearly independent. Since $\mathcal{B}$ is a linearly independent set of n vectors, it must be a basis for $\mathbb{R}^n$.

P11.9 Let A and B be orthogonal matrices. Then, by definition, $A^{-1} = A^\top$, and $B^{-1} = B^\top$. Now

$$\begin{aligned} (AB)^\top AB &= B^\top A^\top AB \\ &= B^{-1}A^{-1}AB \\ &= B^{-1}IB \\ &= I. \end{aligned}$$

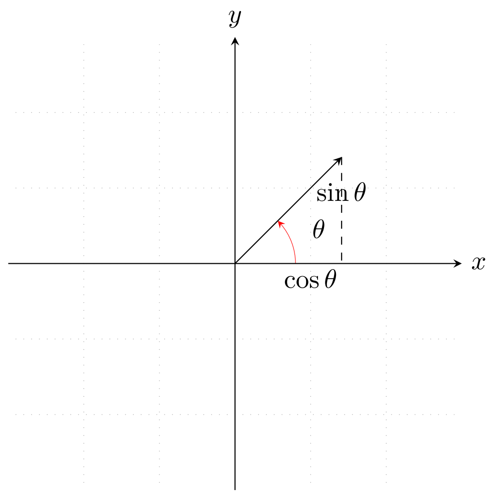
We conclude that AB is invertible and $(AB)^{-1} = (AB)^\top$. Therefore, AB is also orthogonal.

P11.10 Let Q be an orthogonal matrix. Then $Q^{-1} = Q^\top$. Note that $\det(Q^\top) = \det(Q)$ by Proposition 7.23 and $\det(Q^{-1}) = \frac{1}{\det(Q)}$ by an Exercise in chapter 7. Combining these, we get

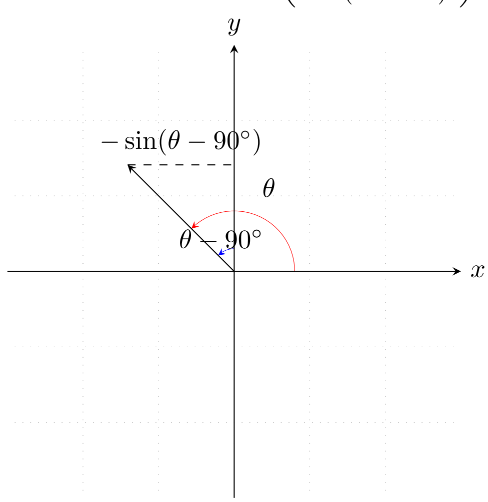
$$\frac{1}{\det(Q)} = \det(Q), \text{ which is equivalent to } (\det(Q))^2 = 1.$$

It follows that $\det(Q) = \pm\sqrt{1} = \pm 1$.

P11.11 (a) Let $\vec{v} \in \mathbb{R}^2$. If $\vec{v}$ is of the form $\vec{v} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$, then $\|\vec{v}\| = \sqrt{\cos^2\theta + \sin^2\theta} = \sqrt{1} = 1$. For the converse, suppose that $\|\vec{v}\| = 1$, and that θ is the angle $\vec{v}$ makes with the positive x-axis. Let's split this up into 4 cases depending on which quadrant $\vec{v}$ is in.
Case 1: $\vec{v}$ is in the first quadrant (both x and y coordinates are positive). In this case, We have the following plot, and using trigonometry, we find

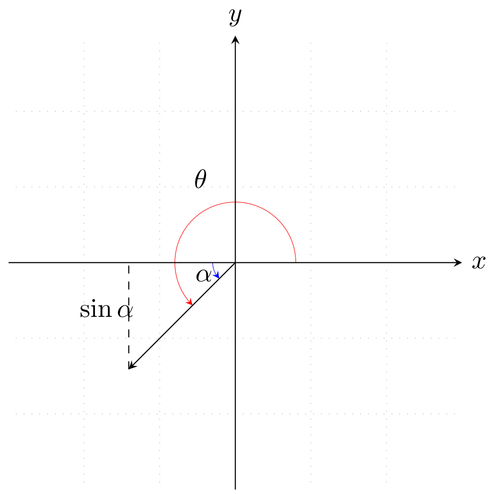


that $\vec{v} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$.
Case 2: $\vec{v}$ is in the second quadrant (the x coordinate is negative and the y coordinate is positive). In this case, We have the following plot, and using trigonometry, we find that $\vec{v} = \begin{pmatrix} -\sin(\theta - 90^\circ) \\ \cos(\theta - 90^\circ) \end{pmatrix} = \begin{pmatrix} \sin(90^\circ - \theta) \\ \cos(90^\circ - \theta) \end{pmatrix} =$



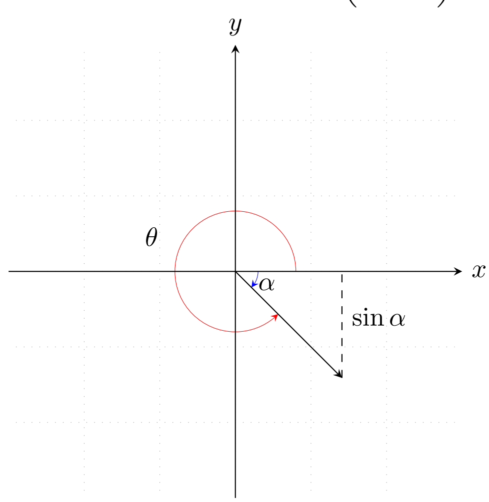
$\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$.
Case 3: $\vec{v}$ is in the third quadrant (both x and y coordinates are negative). In this case, we have the following plot (where $\alpha = \theta - 180^\circ$), and

using trigonometry, we find that $\vec{v} = \begin{pmatrix} -\cos\alpha \\ -\sin\alpha \end{pmatrix} = \begin{pmatrix} -\cos(\theta-180°) \\ -\sin(\theta-180°) \end{pmatrix} =$



$\begin{pmatrix} -\cos(180°-\theta) \\ \sin(180°-\theta) \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$.

Case 4: $\vec{v}$ is in the fourth quadrant (the x coordinate is positive and the y coordinate is negative). In this case, We have the following plot (where $\alpha = 360° - \theta$), and using trigonometry, we find that $\vec{v} = \begin{pmatrix} \cos\alpha \\ -\sin\alpha \end{pmatrix} =$



$\begin{pmatrix} \cos(360°-\theta) \\ -\sin(360°-\theta) \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$.

(b) Let Q be an orthogonal 2×2 matrix. Then the columns of Q form an orthonormal basis for $\mathbb{R}^2$. Let $\vec{v}$ be the first column of Q, then $\|\vec{v}\| = 1$ and using the first part, we have $\vec{v}$ is of the form $\vec{v} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$. If $\vec{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ is the second column of Q, then $\vec{u}\cdot\vec{v} = 0$, which gives us $u_1\cos\theta + u_2\sin\theta = 0$. It follows that $\vec{u} \in \text{Span}\begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$, so $\vec{u} = c\begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$ for some scalar c. Then $\|\vec{u}\|^2 = u_1^2 + u_2^2 = (-csin\theta)^2 + (c\cos\theta)^2 = c^2(sin^2\theta + \cos^2\theta) = c^2$, or $\|\vec{u}\| = |c|$. Since $\vec{u}$ is part of an orthonormal basis, we have $\|\vec{u}\| = 1$, so that $c = \pm 1$, and we get that $\vec{u} = \pm \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$. Therefore, we have

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ or } Q = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

(c)

P11.12 Let $S = \{\vec{v}_1, \ldots, \vec{v}_m\}$ be a set of nonzero orthogonal vectors in $\mathbb{R}^n$. We will show that S is linearly independent. Consider the vector equation

$$a_1\vec{v}_1 + \cdots + a_m\vec{v}_m = \vec{0}.$$

Take the dot product of both sides of this equation with $\vec{v}_i$, where $i \in \{1, \cdots, m\}$. We will get

$$a_1\vec{v}_1 \cdot \vec{v}_i + \cdots + a_m\vec{v}_m \cdot \vec{v}_i = \vec{0} \cdot \vec{v}_i = 0.$$

Since the vectors are orthogonal, $\vec{v}_j \cdot \vec{v}_i = 0$ whenever $i \neq j$, and the last equation reduces to

$$a_i\vec{v}_i \cdot \vec{v}_i = 0.$$

In other words, we have $a_i\|\vec{v}_i\|^2 = 0$. We assumed that $\vec{v}_i$ is a nonzero vector, and so by a previous exercise, $\|\vec{v}_i\|^2 \neq 0$. It follows that $a_i = 0$ for each $i \in \{1, \cdots, m\}$, and therefore, the set S is linearly independent.

P11.13 We first prove a claim that if A is any matrix and B is invertible then $\operatorname{rank}(AB) = \operatorname{rank}(A)$, and $\operatorname{rank}(BA) = \operatorname{rank}(A)$, whenever BA or AB are defined. Note that the column space of AB is the same as the column space of A. If $\vec{v} \in \operatorname{Col}(A)$, then $A\vec{x} = \vec{v}$ for some vector $\vec{x}$, and in this case, $AB(B^{-1}\vec{x}) = A\vec{x} = \vec{v}$, implying that $\vec{v} \in \operatorname{Col}(AB)$. If $\vec{v} \in \operatorname{Col}(AB)$, then $AB\vec{x} = \vec{v}$ for some vector $\vec{x}$, but then $A(B\vec{x}) = AB\vec{x} = \vec{v}$, so that $\vec{v} \in \operatorname{Col}(A)$. Since AB and A have the same column space, they must have the same rank (recall that the rank of a matrix is the dimension of its column space.) Now to prove the second part of the claim. Note that BA and A have the same null space. If $\vec{x} \in \operatorname{Nul}(A)$, then $A\vec{x} = \vec{0}$, and $BA\vec{x} = B\vec{0} = \vec{0}$ so that $\vec{x} \in \operatorname{Nul}(BA)$. Also, if $\vec{x} \in \operatorname{Nul}(BA)$, then $BA\vec{x} = \vec{0}$, multiplying by B^{-1}, we get $A\vec{x} = B^{-1}\vec{0} = \vec{0}$, so that $\vec{x} \in \operatorname{Nul}(A)$. Therefore, $\operatorname{Nul}(A) = \operatorname{Nul}(BA)$, so the nullity of A is equal to the nullity of BA. By the rank and nullity theorem, it follows that A and B have the same rank.

Now let's get back to the original problem. Let A be a matrix with singular value decomposition $A = U\Sigma V^\top$. Then $\operatorname{rank}(A) = \operatorname{rank}(U\Sigma V^\top)$. Since U and $V^\top$ are invertible matrices, it follows that $\operatorname{rank}(A) = \operatorname{rank}(\Sigma)$. But Σ is a (possibly non-square) diagonal matrix, and so its rank is just the number of its non-zero diagonal entries, which is exactly the number of the nonzero singular values of A.

P11.14 Let A be a square matrix with singular value decomposition $A = U\Sigma V^\top$. Then $\operatorname{rank}(A) = \operatorname{rank}(U\Sigma V^\top)$. Then $\det(A) = \det(U\Sigma V^\top) = \det(U)\det(\Sigma)\det(V^\top)$. We know that U and V are invertible matrices, and so $\det(U) \neq 0$, and $\det(V) = \det(V^\top) \neq 0$. Therefore, $\det(A) = 0$ if and only if $\det(\Sigma) = 0$. And $\det(\Sigma)$ is the product of its diagonal entries, which are the singular values of A. Therefore, $\det(A) \neq 0$ and A is invertible if and only if 0 is not a singular value of A.

P11.15 Let A be an $m \times n$ matrix with singular value decomposition $A = U\Sigma V^\top$. Then $A^\top = (U\Sigma V^\top)^\top = V\Sigma^\top U^\top$. Note that this is a singular value decomposition of $A^\top$, since U is an $m \times m$ orthogonal matrix implying that $U^\top$ is also an orthogonal matrix, and V is an $n \times n$ orthogonal matrix, implying that $V^\top$ is also an orthogonal matrix. Now Σ is a block diagonal matrix, and so its transpose is also block diagonal and has the same non-zero singular values on the diagonal. Therefore, $A^\top$ has the same non-zero singular values as A.

Appendix D

AMPA Toolkit

The mathematical process involves three distinct and equally important steps: (1) *solve* a problem or develop a theory, (2) *write* a rigorous solution that can be formally verified, and (3) *communicate* your argument in a way that is both rigorous and understandable. In tutorial and class activities, we have been practicing steps (2) and (3) together. In this Appendix, we outline a few strategies to help with step (1). It's important to note that this step is often messy – while mathematics is often presented in it's cleanest version, that almost never represents what needed to happen behind the scenes.

Problem solving is a skill that takes time and practice to develop, and is really a process of determination and creativity. With that said, there are a few concrete tools that we can give to help you along in this process. The following tools were adapted from the "AMPA Framework" ([**1**]). This is a framework that mathematics education researchers developed to characterize the types of activities mathematicians engage in during their authentic professional lives.

AMPA Toolkit. When encountering a problem you're unsure how to solve, try the following.

(1) Generate Examples. Generating examples can help us test hypotheses and detect whether a statement is true or false.

(2) Create Diagrams. Sometimes we can give a visual representation of an object to help us capture its structural features. Even if your drawing isn't accurate, it can help with the 'big picture' understanding of the problem.

(3) Make Conjectures. When we don't know where to go, it's often useful to make an educated guess (what mathematicians call a conjecture) and then test that conjecture. If the conjecture turns out to be false, refine it and try again.

(4) Formalize and Deformalize. Sometimes it's helpful to change the formality of a statement. This can go both directions: oftentimes it's helpful to recall formal definitions and work with the problem symbolically. Other times, it can be helpful to deconstruct a formal statement into natural language to help us gain intuition. A good example of this is our two definitions of linear dependence: the geometric definition is more intuitive (which can help us gain understanding), while the algebraic definition is more formal (making it easier to work with symbolically). Depending on the problem, working with one or the other, or moving back and forth between the two, may be helpful.

Bibliography

[1] Kathleen Melhuish, Kristen Vroom, Kristen Lew, and Brittney Ellis. Operationalizing authentic mathematical proof activity using disciplinary tools. *The Journal of Mathematical Behavior*, 68:101009, 2022.

[2] D Poole. *Linear Algebra: A Modern Introduction*. Thomson Brooks/Cole, 2006.

[3] Gilbert Strang. *Introduction to linear algebra*. SIAM, 2022.